

SIMULATING THE COSMIC DISTRIBUTION OF  
NEUTRAL HYDROGEN  
AND  
ITS CONNECTION WITH GALAXIES

ALIREZA RAHMATI

ISBN 978-9-46-191899-4

Cover by A. Rahmati

**Front:** The column density distribution of neutral hydrogen in a photoshoped simulated galaxy. This design is motivated by the main findings of Chapter 4.

**Back:** Ancient constellations as published in *The book of fixed stars* by the famous Iranian astronomer, Abd al-Rahman al-Sufi in 964 A.D. This highly influential book has been reproduced numerous times during the last millennium. The illustrations used for the design of the back-cover are taken from a version produced in 1430-1440 A.D. in Samarkand (Uzbekistan), which is accessible through the following link:

<http://gallica.bnf.fr/ark:/12148/btv1b60006156>

SIMULATING THE COSMIC DISTRIBUTION OF  
NEUTRAL HYDROGEN  
AND  
ITS CONNECTION WITH GALAXIES

PROEFSCHRIFT

ter verkrijging van  
de graad van Doctor aan de Universiteit Leiden,  
op gezag van de Rector Magnificus prof. mr. C. J. J. M. Stolker,  
volgens besluit van het College voor Promoties  
te verdedigen op dinsdag 15 oktober 2013  
klokke 11:15 uur

door  
Alireza Rahmati  
geboren te Qom, Iran  
in 1982

Promotiecommissie

Promotor: Prof. dr. J. Schaye

Overige leden: Prof. dr. M. Franx  
Dr. A. H. Pawlik (Max-Planck Institute for Astrophysics)  
Prof. dr. S. F. Portegies Zwart  
Prof. dr. J. X. Prochaska (University of California, Santa Cruz)  
Prof. dr. H. J. A. Röttgering  
Prof. dr. P. P. van der Werf

# TABLE OF CONTENTS

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Current standard model for galaxy formation and evolution . . . .	2
1.1.1	Cosmology: the backbone of galaxy formation . . . . .	2
1.1.2	Main physical processes that drive galaxy evolution . . . .	3
1.2	Neutral hydrogen in galactic ecosystems . . . . .	4
1.3	Simulations . . . . .	6
1.3.1	Cosmological hydrodynamical simulations . . . . .	6
1.3.2	Radiative transfer . . . . .	8
1.3.3	Radiative transfer with TRAPHIC . . . . .	10
1.4	This thesis . . . . .	11
	References . . . . .	13
<b>2</b>	<b>On the evolution of the H I column density distribution in cosmological simulations</b>	<b>15</b>
2.1	Introduction . . . . .	16
2.2	Simulation techniques . . . . .	18
2.2.1	Hydrodynamical simulations . . . . .	18
2.2.2	Radiative transfer with TRAPHIC . . . . .	19
2.2.3	Ionizing background radiation . . . . .	22
2.2.4	Recombination radiation . . . . .	26
2.2.5	The H I column density distribution function . . . . .	26
2.2.6	Dust and molecular hydrogen . . . . .	27
2.3	Results . . . . .	30
2.3.1	Comparison with observations . . . . .	30
2.3.2	The shape of the H I CDDF . . . . .	34
2.3.3	Photoionization rate as a function of density . . . . .	36
2.3.4	The roles of diffuse recombination radiation and collisional ionization at $z = 3$ . . . . .	39
2.3.5	Evolution . . . . .	41
2.4	Conclusions . . . . .	45
	Acknowledgments . . . . .	46
	References . . . . .	47
	Appendix A: Photoionization rate as a function of density . . . . .	49
	A1: Replacing the RT simulations with a fitting function . . . . .	49
	A2: The equilibrium hydrogen neutral fraction . . . . .	50
	Appendix B: The effects of box size, cosmological parameters and resolution on the H I CDDF . . . . .	53
	Appendix C: RT convergence tests . . . . .	53
	C1: Angular resolution . . . . .	53
	C2: The number of ViP neighbors . . . . .	53
	C3: Direct comparison with another RT method . . . . .	55
	Appendix D: Approximated processes . . . . .	56
	D1: Multifrequency effects . . . . .	56

## TABLE OF CONTENTS

---

D2: Helium treatment . . . . .	57
<b>3 The impact of local stellar radiation on the H<sub>I</sub> column density distribution</b>	<b>59</b>
3.1 Introduction . . . . .	61
3.2 Photoionization rate in star-forming regions . . . . .	63
3.3 Simulation techniques . . . . .	64
3.3.1 Hydrodynamical simulations . . . . .	64
3.3.2 Radiative transfer . . . . .	65
3.3.3 Ionizing background radiation and diffuse recombination radiation . . . . .	67
3.3.4 Stellar ionizing radiation . . . . .	68
3.4 Results and discussion . . . . .	70
3.4.1 The role of local stellar radiation in hydrogen ionization . . . . .	70
3.4.2 Star-forming particles versus stellar particles . . . . .	76
3.4.3 Stellar ionizing radiation, its escape fraction and the buildup of the UVB . . . . .	79
3.4.4 The impact of local stellar radiation on the H <sub>I</sub> column density distribution . . . . .	82
3.5 Discussion and conclusions . . . . .	87
Acknowledgments . . . . .	91
References . . . . .	91
Appendix A: Hydrogen molecular fraction . . . . .	93
Appendix B: Resolution effects . . . . .	94
B1: Limited spatial resolution at high densities . . . . .	94
B2: The impact of a higher resolution on the RT . . . . .	95
Appendix C: Calculation of the escape fraction . . . . .	97
<b>4 Predictions for the relation between strong H<sub>I</sub> absorbers and galaxies at redshift 3</b>	<b>99</b>
4.1 Introduction . . . . .	100
4.2 Simulation techniques . . . . .	102
4.2.1 Hydrodynamical simulations . . . . .	102
4.2.2 Finding galaxies . . . . .	103
4.2.3 Finding strong H <sub>I</sub> absorbers . . . . .	103
4.2.4 Connecting H <sub>I</sub> absorbers to galaxies . . . . .	107
4.3 Results and discussion . . . . .	108
4.3.1 Spatial distribution of H <sub>I</sub> absorbers . . . . .	110
4.3.2 The effect of a finite detection threshold . . . . .	111
4.3.3 Distribution of H <sub>I</sub> absorbers relative to halos . . . . .	114
4.3.4 Resolution limit in simulations . . . . .	115
4.3.5 Correlations between absorbers and various properties of their associated galaxies . . . . .	119

---

4.3.6	Are most strong HI absorbers at $z \sim 3$ around Lyman-Break galaxies? . . . . .	124
4.4	Summary and conclusions . . . . .	126
	Acknowledgments . . . . .	128
	References . . . . .	128
	Appendix A: Choosing the maximum allowed LOS Velocity difference . . . . .	131
	Appendix B: Impact of feedback . . . . .	131
	Appendix C: Impact of local stellar radiation . . . . .	134
	Appendix D: Resolution tests . . . . .	135
<b>5</b>	<b>Genesis of the dusty Universe: modeling submillimetre source counts</b>	<b>141</b>
5.1	Introduction . . . . .	143
5.2	Model ingredients . . . . .	145
5.2.1	CLF at $z = 0$ . . . . .	146
5.2.2	CLF evolution . . . . .	147
5.2.3	SED Model . . . . .	149
5.2.4	The algorithm . . . . .	150
5.3	850 $\mu\text{m}$ Observational Constraints . . . . .	152
5.3.1	Observed 850 $\mu\text{m}$ source count . . . . .	152
5.3.2	Redshift distribution of bright 850 $\mu\text{m}$ sources . . . . .	153
5.4	Finding 850 $\mu\text{m}$ best-fit Model . . . . .	154
5.4.1	The source count curve: Amplitude vs. Shape . . . . .	154
5.4.2	The luminosity evolution . . . . .	155
5.5	Other necessary model ingredients . . . . .	157
5.5.1	The 850 $\mu\text{m}$ best-fit model . . . . .	159
5.6	Other wavelengths . . . . .	160
5.6.1	Long submm wavelengths: 850 $\mu\text{m}$ and 1100 $\mu\text{m}$ . . . . .	162
5.6.2	SPIRE intermediate wavelengths: 500 $\mu\text{m}$ , 350 $\mu\text{m}$ and 250 $\mu\text{m}$ . . . . .	164
5.6.3	Short wavelengths: 160 $\mu\text{m}$ and 70 $\mu\text{m}$ . . . . .	165
5.6.4	A best-fit model for all wavelengths . . . . .	166
5.7	Discussion . . . . .	171
5.7.1	The implied evolution scenario for dusty galaxies . . . . .	171
5.7.2	Our best-fit model and previous models . . . . .	175
5.8	Conclusions . . . . .	177
	Acknowledgments . . . . .	178
	References . . . . .	179
	Appendix A: Some numerical details . . . . .	181
	<b>Nederlandse samenvatting</b>	<b>183</b>
	<b>Publications</b>	<b>189</b>
	<b>Curriculum Vitae</b>	<b>191</b>

---





# 1

## INTRODUCTION

For more than a millennium, we have been aware of the existence of a celestial small cloud in the constellation of Andromeda (al-Sufi, 964), but it has been less than a century since we observed that the small cloud, which is commonly known to us as M31, is a spiral galaxy outside of our own Milky Way. In fact, almost everything we know about M31 and other galaxies has been learned during the last century, thanks to the advent of large telescopes and new technologies. During the last few decades, theoretical models have helped us to make sense of the rapidly increasing wealth of data and have caused a revolution in our understanding of the Universe. In particular, as a result of advancements in numerical techniques and computational recourses, cosmological simulations have become an indispensable tool for probing the main processes involved in galaxy formation and evolution.

This thesis is an attempt to add to our understanding of the Universe, by using cosmological simulations for studying the distribution and evolution of the neutral hydrogen around galaxies. In this work, I extensively use state-of-the-art cosmological hydrodynamical simulations of galaxy formation. To set the stage, I begin this chapter with a very brief overview of our current knowledge about how galaxies form and evolve. I continue by explaining the importance of studying the neutral hydrogen for understanding galaxies. Then, I discuss briefly how hydrodynamical and radiative transfer simulations work, before ending this introductory chapter with the outline of this thesis.

## 1.1 Current standard model for galaxy formation and evolution

The theory of galaxy formation is where the properties of the largest scales in the Universe meet the physics at tiny atomic scales. The huge dynamic range of the relevant scales spanned by the different physical entities (e.g., length, mass, time) that characterize galaxies, makes it a daunting task to model these complex systems. Despite the monstrous size of this problem, in principle, the formation and evolution of galaxies should be understandable in terms of the known physical laws. Using these physical laws to explain the observed trends among galaxies has been the main challenge of the theory of galaxy formation and evolution. As a result of the hard work of many great minds, we now have an understanding of how galaxies form and evolve, which can explain, with a good degree of accuracy, what we see in the Universe. However, as in other natural sciences, any progress in understanding galaxies reveals new puzzles to solve. As a result, there is (and there will be) a huge number of phenomena that we do not fully understand. In the following, I briefly review what I call the *standard* model of galaxy formation and evolution, which consists of ideas that glue together large sets of observed properties, and are commonly accepted by most experts who work in this field.

### 1.1.1 Cosmology: the backbone of galaxy formation

Cosmology serves as the backbone of galaxy formation by describing the physical laws that govern the formation and evolution of structures on large scales. In this context, the  $\Lambda$ CDM paradigm has enjoyed great success by explaining a large number of observables, such as the temperature fluctuations in the radiation we receive from the early Universe (i.e., the Cosmic Microwave Background), the accelerated expansion of the Universe and the growth of structures. The constraints on the parameters of the  $\Lambda$ CDM concordance cosmological model have been improving rapidly in recent years, thanks to precise observational experiments like the Cosmic Background Explorer (COBE: Mather et al., 1990), the Wilkinson Microwave Anisotropy Probe (WMAP: Bennett et al., 2003; Spergel et al., 2003; Komatsu et al., 2011) and recently, the Planck satellite (Planck Collaboration et al., 2013). Based on the above mentioned measurements, and other experiments, we know that the vast majority of the energy density of the Universe is in the form of Dark Energy (i.e.,  $\Lambda$ ), which causes the accelerating expansion of the Universe. The rest consists mainly of cold dark matter (i.e., CDM) with tiny fractions of baryonic matter and radiation. The exact values of the above mentioned fractions define the dynamics of the Universe on large scales and the growth of small scale density fluctuations in the nearly homogeneous initial distribution of (Dark and baryonic) matter after the Big Bang.

As the Universe expands, the separation between small scale density fluctuations increases. At the same time, the deviation between the density of fluctuations and the mean density of the Universe also increases. In other words, over-dense regions attract more matter at the expense of draining under-dense regions. When the density fluctuation (i.e.,  $\Delta\rho$ ) is much smaller than the mean density of the Universe (i.e.,  $\Delta\rho/\rho \ll 1$ ), its physical size increases with the expansion of the Universe as the magnitude of  $\Delta\rho$  increases (i.e., the linear regime). There is, however, a turn-around when  $\Delta\rho/\rho \sim 1$ , after which the physical size of the fluctuation starts to decrease (i.e., the non-linear regime). The result of the latter stage is the formation of self-gravitating bound structures. Since the majority of matter is in the form of collisionless dark matter, collapsing structures (i.e., dark matter haloes) are regularized through violent relaxation. On large scales and low over-densities, the baryonic matter is bound to the dark matter, which is the dominant gravitational actor. In the non-linear regime, however, baryons reveal their collisional nature and become shock heated to very high temperatures through the efficient conversion of gravitational energy into the internal energy of the baryonic gas, as it collapses into the inner parts of the gravitational potential wells. At this stage, the fate of baryons in the collapsed structures is significantly affected by the laws of small-scale atomic physics and related complex and collective processes like radiative cooling and star formation. At the same time, the large-scale evolution of structures continues and brings more matter into the collapsed structures and merges them.

### 1.1.2 Main physical processes that drive galaxy evolution

Despite the great success of the  $\Lambda$ CDM paradigm in explaining the invisible side of galaxy formation by accurately predicting the formation and growth of dark matter haloes (e.g., Zehavi et al., 2011; Heymans et al., 2013), the complex physical processes that control the evolution of gas and stars are far from understood. To first order, the baryonic content of collapsing dark matter haloes, which is initially shock heated to high temperatures, loses its energy through radiative cooling. The conservation of angular momentum dictates the cooling gas to form a rotating disk as it loses energy and falls deeper into the potential well of the dark matter halo. The density of baryonic gas increases until the radiative cooling becomes inefficient and the gaseous disk approaches a quasi-equilibrium state. The small scale instabilities in the gaseous disk, however, continue to grow into dense molecular clouds which evolve and collapse to form stars.

The advent of stars makes the lives of galaxies much more complicated. Stars are energetic sources of feedback and significantly affect the fate of their parent galaxies. They heat up the gas and ionize it with their radiation, and they produce heavy elements as they evolve. These processes change the evolution of the gas around stars by changing its cooling/heating. Stars also transport large quantities of kinetic energy and heavy elements into their surroundings by injecting winds into the interstellar medium (ISM). Massive stars, which evolve

faster than stars with lower masses, end their lives dramatically in energetic explosions (i.e., Supernovae; SNe) that inject huge amounts of energy into the ISM of their host galaxies, affecting subsequent star formation and launching large-scale galactic winds. In addition to energetic SNe explosions, radiation pressure from very luminous young stars removes gas from star-forming regions, and possibly makes a significant contribution to the launching of galactic winds.

The presence of very massive black holes, the end result of the death of very massive stars, causes more complications in our understanding of how galaxies evolve. Super-massive black holes, which are found at the centers of many (if not all) galaxies we see in the Universe, are very efficient in converting mass into energy. The huge amount of energy they inject into their host galaxies in different ways, changes their fate dramatically.

There is much solid observational evidence and there are strong theoretical arguments for the presence and importance of the above mentioned mechanisms in the formation and evolution of galaxies. Understanding how they work individually and together to control the lives of galaxies at different epochs has been an active area of research during the last few decades, an area of research which is still vibrantly active due to its complexity.

## 1.2 Neutral hydrogen in galactic ecosystems

As mentioned above, galaxies are influenced on the one hand by the force of gravity, which forms haloes, brings fresh material into the already existing haloes and keeps baryonic and dark matter structures together, and on the other hand, by the feedback mechanisms that fight against the force of gravity and try to unbind structures. The interaction between these two fronts creates a complex ecosystem in and around galaxies, and imprints the history of galaxies in the distribution of baryons around them (i.e., the circumgalactic medium; CGM). In this context, understanding the distribution of neutral hydrogen (HI) is of particular importance. The main reason for this is that HI is the main fuel for the formation of molecular clouds, the birth places of stars, which makes studying the distribution of HI and its evolution crucial for our understanding of various aspects of star formation.

We know that shortly after the Big Bang, the initially hot plasma cools down as the Universe expands and electrons and protons recombine. Hydrogen, which is the most abundant element in the Universe, becomes highly ionized again by  $z \sim 6$  due to the formation of the first stars and galaxies (i.e., the reionization). After reionization, the mean-free-path of ionizing photons (i.e., the distance photons can travel before being significantly absorbed) increases with time, as the average star formation activity of the Universe increases and the Universe becomes less dense. The relatively uniform distribution of sources of ionizing radiation on large scales creates a relatively uniform ultraviolet background radiation (i.e., the UVB) which is dominated by stellar photons at  $z \gtrsim 3$

and quasars at lower redshifts (e.g., Becker & Bolton, 2013). As a result, after reionization most hydrogen atoms are kept ionized by the UVB radiation they receive from all the ionizing sources they can see in the Universe.

The complexity of processes that set the ionization state of hydrogen depends on the HI column density. At low HI column densities (i.e.,  $N_{\text{HI}} \lesssim 10^{17} \text{ cm}^{-2}$ , corresponding to the so-called Lyman- $\alpha$  forest), hydrogen is highly ionized by the UVB radiation and largely transparent to the ionizing radiation. For these systems, the HI column densities can therefore be accurately computed in the optically thin limit. At higher HI column densities (i.e.,  $N_{\text{HI}} \gtrsim 10^{17} \text{ cm}^{-2}$ , corresponding to the so-called Lyman Limit and Damped Lyman- $\alpha$  systems), the gas becomes optically thick and self-shielded. As a result, the ionization state of hydrogen in these systems is more sensitive to various radiative transfer effects such as self-shielding, shadowing and the fluctuations of the UVB radiation on small scales.

Studying the distribution of HI is proven to be challenging. In the local Universe, the HI content of galaxies can be probed by observing 21-cm emission, but at higher redshifts this will not be possible until the advent of significantly more powerful telescopes, such as the Square Kilometer Array<sup>1</sup>. At  $z \lesssim 6$ , i.e., after reionization, the neutral gas can be probed through absorption signatures that are imprinted by the intervening HI systems on the spectra of bright background sources, such as quasars. The analysis of these absorption features provides an alternative probe of the distribution of matter at high redshifts, compared to studying the Universe through emission. The large distances that separate most absorbers from their background QSOs make it unlikely that there is a physical connection between them. This opens up a window to study an unbiased sample of matter that resides between us and the background QSOs.

Constraining the statistical properties of the HI distribution has been the focus of many observational studies during the last few decades (e.g., Tytler, 1987; Kim et al., 2002; Péroux et al., 2005; O’Meara et al., 2007; Noterdaeme et al., 2009; Prochaska et al., 2009; O’Meara et al., 2013). Thanks to a significant increase in the number of observed quasars and improved observational techniques, more recent studies have extended these observations to both lower and higher HI column densities and to higher redshifts. Therefore, it is important to study the properties of the HI distribution in cosmological simulations to better understand the observed trends and to put them in the context of the standard theory of galaxy formation and evolution, which is the focus of this thesis. Once simulations agree with observations, one can use them to predict what future observations reveal. These predictions can be used to validate the underlying models in the simulations and to examine the importance of different processes that are relevant to the formation and evolution of galaxies, or in case of disagreement, to point us to necessary improvements.

---

<sup>1</sup><http://www.skatelescope.org/>

## 1.3 Simulations

Hydrodynamical simulations attempt to model complex baryonic interactions by combining various physically motivated and empirical ingredients. Modern state-of-the-art cosmological simulations of galaxy formation try to put together most of what we know about the astrophysical and cosmological processes that are shaping the Universe on different scales and they try to reproduce different observables.

In this context, the vital importance of radiation and radiative transfer (RT) processes for galaxy evolution in general, and for producing the HI distribution in particular, is evident. However, RT is often ignored or poorly approximated in the simulations. The main reason for this is the high dimensionality of the underlying calculation which makes RT an enormous computational challenge for cosmological simulations. Thanks to increasingly more powerful computers, and in the light of new algorithms, the accurate treatment of radiative effects is now becoming possible.

Since in this thesis we extensively use hydrodynamical simulations and RT, we will briefly discuss how they work.

### 1.3.1 Cosmological hydrodynamical simulations

Cosmological simulations calculate the evolution of the Universe by starting from an approximately uniform density of matter with small fluctuations on different scales. These fluctuations are set by the observed statistical properties of the fluctuations in the early stages of the Universe (i.e., the CMB). To be able to trace the evolution of the Universe numerically, different techniques are adopted to discretize the continuous distribution of matter into a finite number of resolution elements (e.g., particles). In addition, to keep the numerical calculations tractable, one needs to adopt a finite volume which is assumed to be a representative sample of the whole Universe. Periodic boundary conditions can then be used which assume that the distribution of matter on scales beyond the extent of the simulation box is statistically similar to that inside it.

The gravitational interactions between all particles in the simulation are followed directly while the large-scale cosmological expansion of the Universe is accounted for by a change of coordinate system. Different techniques are used to accelerate the computation of the gravitational field without losing accuracy significantly. For instance, calculating the gravitational interaction between two groups of particles that are far away from each other (compared to the typical distances between the particles in each group) is possible even if we neglect the small scale distribution of the group members and replace the whole group with a single gravitationally interacting element (see for an example the TreePM algorithm explained in [Bagla, 2002](#)).

Calculating the hydrodynamical forces, which are short-range forces compared to gravity, is simpler due to their local nature. In other words, only neigh-

---

boring resolution elements interact hydrodynamically. One of the most successful methods to simulate the evolution of the gas fluid is the smoothed particle hydrodynamic (SPH) technique which was introduced by [Gingold & Monaghan \(1977\)](#) and [Lucy \(1977\)](#). In the SPH prescription, the fluid is discretized into individual particles that are smoothed. This means that the relevant properties of the fluid that are represented by each particle are distributed smoothly in space around the particle. Then, the value of each relevant quantity (in most cases only density), for any given position, is calculated by adding the contribution of all SPH particles that are close enough to have significant contributions. Then, the evolution of the fluid is calculated based on the known physical laws that affect the fluid, like the laws of mass, momentum and energy conservation.

By combining the gravitational and hydrodynamical forces, it is possible to start from small initial fluctuations and trace the evolution of dark matter and baryonic gas as the Universe evolves. As the formation of structure proceeds, the baryonic content of collapsing structures, which is initially shock heated, cools down due to radiative cooling. The radiative cooling is calculated in the simulations by including several important atomic processes. These processes are mainly sensitive to the density, temperature, the abundances of different elements (i.e., metallicity) and the properties of the radiation field. Because of the high-dimensionality of the equations that control the cooling, their net effect is included in the simulations using fitting functions and tables. Due to cooling, the density of baryons increases which further complicates the simulation of the evolution of baryons.

The limited spatial resolution of cosmological simulations, which is usually of the order of a kilo-parsec, is not enough to resolve the complex and multi-phase structure of the ISM. Because of this, an effective equation of state is often adopted to model the collective hydrodynamical properties of the ISM gas. This effective equation of state can also be used to prevent artificial fragmentation of dense gas on scales close to the resolution limit ([Schaye & Dalla Vecchia, 2008](#)). As the gas density increases, the dense ISM should be converted into stars. Since it is not possible to simulate the complex process of star formation at kpc resolution, simplified algorithms are used to convert gas into stars. These star formation prescriptions are often tuned to match the observed relation between the gas (column) density and star formation rate in real galaxies, on kpc scales (i.e., the [Kennicutt-Schmidt relation](#); [Kennicutt, 1998](#)).

The resolution elements (e.g., SPH particles) that satisfy the adopted star formation criteria are converted into stellar particles with typical masses of  $10^5 - 10^6 M_{\odot}$ . In other words, each stellar particle in cosmological simulations represent groups of stars which are comparable to observed stellar clusters. The subsequent evolution of this population of stars, which are assumed to be formed at the same time, is followed by assuming an initial mass function (IMF) and using stellar population synthesis models that trace the stellar evolution. As the simulation continues, the stars that are assumed to be embedded in each stellar particle, age and produce metals. The fraction of the stars that are massive



enough to explode as SNe, end their lives in dramatic explosions and heat up the gas which also launches high velocity flows that create large-scale galactic winds. All these processes, together with other complicated phenomena, are poorly understood and their detailed simulation requires much higher resolution than what is affordable in current cosmological simulations. Therefore, they are included in the simulations through approximate rules, similar to the formation of stars explained above, which are commonly known as sub-grid models. The sub-grid models in cosmological simulations try to incorporate our physical knowledge or observed empirical relations into simulations. Due to the large uncertainties, they are often tuned to produce the desired observables, such as the average star formation activity of the Universe or the stellar mass function at different epochs. Because of the complexity of how different subgrid models work individually and in combination, implementing them reliably in the simulations and understanding how they regulate galaxies is an active area of research.

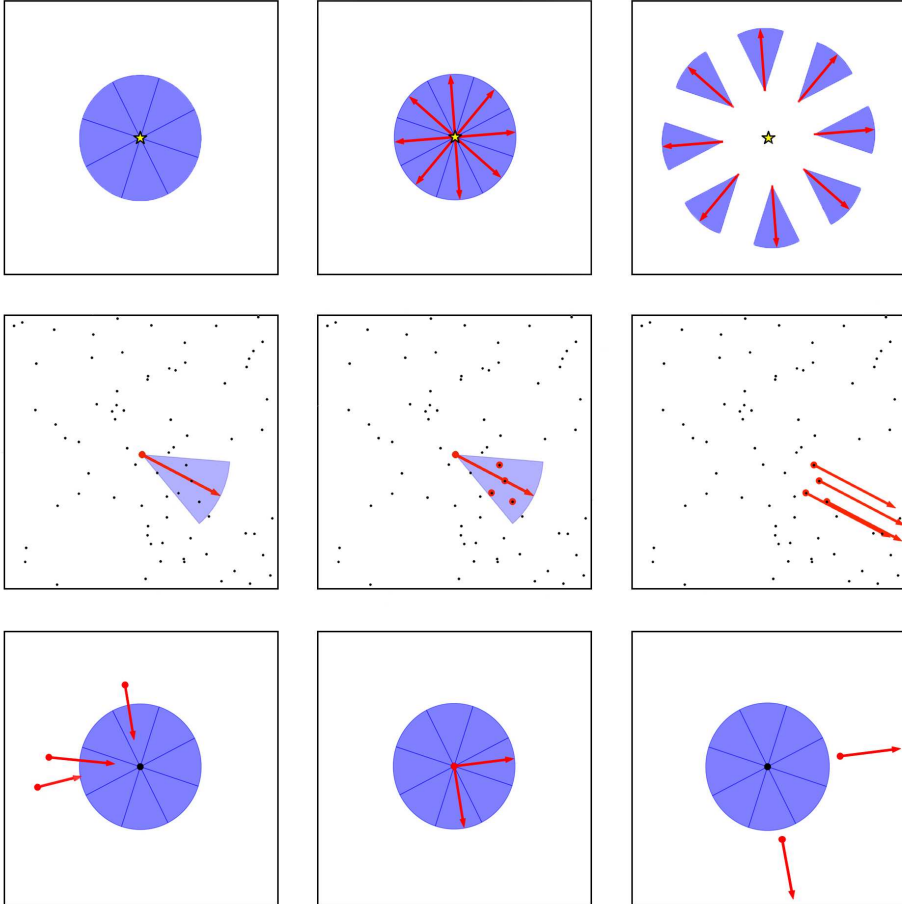
By combining all the above mentioned ingredients (i.e., gravity, hydrodynamics and subgrid models), it is possible to start from initial conditions that are set by cosmological observations, and to simulate the formation and evolution of galaxies and the cosmic distribution of gas.

### **1.3.2 Radiative transfer**

In addition to hydrodynamical and gravitational interactions, radiation plays an important role in shaping the cosmos as we see it. The interaction between light and matter has important consequences for the evolution of structures across a range of different scales. While the radiation pressure close to stars, and perhaps on galactic scales, changes the dynamics of the gas, the UVB radiation, which is the superposition of radiation from large numbers of sources distributed on large scales, affects the cooling/heating of the baryons significantly. In addition, it is impossible to interpret the observations without understanding the relation between the measured intensity of radiation and different phenomena that produce/affect photons.

Although it is essential to include the impact of radiation accurately in cosmological simulations, it is tremendously challenging to follow the production of photons and their propagation as they interact with matter (i.e., radiative transfer). First of all, the details of the RT depend on frequency, which can change as photons travel through space. In addition, photons can travel to large distances which makes them important over a wide range of scales. Because radiation can propagate far from the locations where photons are generated, radiation is more similar to gravity than hydrodynamical interactions. However, unlike gravity, for calculating the radiation field it is important to know what is between the sources of radiation and the point at which the radiation field is calculated. The time-dependency of the radiation field adds further complexity to RT calculations.





**Figure 1.1:** Different steps in the photon transport with TRAPHIC : 1- Emission (top panels): Photons are emitted isotropically from source particles to their SPH neighbors and then travel down-stream based on their propagation directions. 2- Transport (middle panels): photon packets are distributed among the neighboring particles that are within the transmission cone. If there is no neighboring particle inside the emission/transmission cones, virtual particles are created to transport photons along their propagation directions (not shown). 3- Merging (bottom panels): if two or more of the photon packets that a particle receives have close enough propagation directions, they are merged into a single photon packet with appropriately averaged direction.

### 1.3.3 Radiative transfer with TRAPHIC

Because of the reasons mentioned above, RT is sensitive to a large number of variables which makes it computationally expensive. Various techniques have been developed to reduce the computational cost of the RT, mainly by adopting different simplifying assumptions. Among different methods, TRAPHIC (TRANsport of PHotons In Cones; Pawlik & Schaye, 2008, 2011) is a unique RT method with several important advantages, as is explained below, for radiation transport in cosmological simulations that use the SPH prescription. Since we use SPH simulations in this work, TRAPHIC is a natural choice to compute the RT. In the following we briefly review how photon transfer is done with TRAPHIC (see also Figure 1.1).

TRAPHIC is an explicitly photon-conserving RT method designed to transport radiation directly on the irregular distribution of SPH particles. This means that unlike many RT methods that use coarse grids to simplify the calculations, TRAPHIC exploits the full dynamic range that is available based on the underlying SPH simulation. Another challenge that RT methods face is the large number of sources in cosmological simulations. The computational cost of most RT methods is proportional to the number of radiation sources, which poses a big challenge for cosmological RT simulations. TRAPHIC solves this problem by tracing photon packets inside a discrete number of cones which renders the computational cost of the RT independent of the number of radiation sources. The two above mentioned advantages make TRAPHIC particularly well-suited for RT calculation in cosmological density fields with a large dynamic range in densities and large numbers of sources.

The photon transport in TRAPHIC proceeds in two steps: the isotropic emission of photon packets by source particles and their subsequent directed propagation on the irregular distribution of SPH particles. After sources emit photon packets isotropically to their neighbors, the photon packets travel along their propagation directions to neighboring SPH particles which are inside their transmission cones. Transmission cones are regular cones with solid angle  $4\pi/N_{\text{TC}}$  and are centered on the propagation direction. The parameter  $N_{\text{TC}}$  sets the angular resolution of the RT. The transmission cones are defined locally at the transmitting particle, and hence the angular resolution of the RT with TRAPHIC is independent of the distance from the source.

It can happen that transmission cones do not contain any neighboring SPH particles. In this case, additional particles (virtual particles, ViPs) are placed inside the transmission cones to accomplish the photon transport. The ViPs, which enable the particle-to-particle transport of photons along any direction independently of the spatially inhomogeneous distribution of the particles, do not affect the SPH simulation and are deleted after the photon packets have been transferred.

An important feature of the RT with TRAPHIC is the merging of photon packets, which guarantees the independence of the RT computational cost from the

number of sources. Photon packets received by each SPH particle are binned based on their propagation directions in  $N_{\text{RC}}$  reception cones. Then, photon packets with identical frequencies that fall in the same reception cone are merged into a single photon packet with a new direction set by the luminosity-weighted sum of the directions of the original photon packets.

Photon packets are transported along their propagation direction until they reach the distance they are allowed to travel within the RT time step by the finite speed of light. At the end of each time step the ionization states of the particles are updated based on the number of absorbed ionizing photons.

By following the aforementioned steps, TRAPHIC calculates the radiation field in cosmological simulations and the ionization state of different species accurately (see Pawlik & Schaye, 2008, 2011). In this work, we mainly use TRAPHIC to calculate the ionization state of hydrogen by taking into account different sources of radiation.

## 1.4 This thesis

Observational studies are moving rapidly beyond studying only the stellar components of galaxies, towards probing the important but complex processes that affect the distribution of gas in and around galaxies. Hydrodynamical cosmological simulations of galaxy formation are also improving rapidly by including better numerical techniques, more physically motivated and better-understood subgrid models and by using higher resolution. It is important, if not essential, to compare observations and simulations in order to improve our understanding of the observational results and to test/improve the simulations.

The neutral hydrogen distribution and its evolution is closely related to various aspects of star formation. This makes understanding and modeling the HI distribution critically important for studying galaxy evolution. The main focus of this thesis is therefore the study of the cosmic distribution of neutral hydrogen using hydrodynamical cosmological simulations. To do this, we combine hydrodynamical cosmological simulations based on the Overwhelmingly Large Simulations (OWLS; Schaye et al., 2010) with accurate radiative transfer, and account for different photoionizing processes.

In **Chapter 2**, we include the radiative transfer effects of the metagalactic UVB radiation and diffuse recombination radiation for hydrogen ionization at redshifts  $z = 5 - 0$ . We focus on studying HI column densities  $N_{\text{HI}} > 10^{16} \text{ cm}^{-2}$ , where RT effects can be important. By modeling more than 12 billion years of evolution of the HI distribution, we show that the predicted HI column density distribution is in excellent agreement with observations and evolves only weakly from  $z = 5$  to  $z = 0$ . We find that the UVB is the dominant source of HI ionization at  $z \gtrsim 1$ , but that collisional ionization becomes more important at lower redshift, which affects the self-shielding significantly. Based on our simulations, we present fitting functions that can be used to accurately calculate the neutral

hydrogen fractions without RT. Given the difficulty of RT simulations, these fitting functions are particularly useful for the next generation of high resolution cosmological simulations.

Stars typically form at very high column densities, where the gas is self-shielded against the external ionizing radiation. This makes local stellar radiation an important source of ionization at these column densities. However, simulating the effect of ionizing stellar radiation for the large numbers of sources that are typical of cosmological simulations, is extremely challenging. As a result, different studies have found different (and often inconclusive) results regarding the importance of local stellar radiation on the H I distribution. In **Chapter 3**, we tackle this problem by combining cosmological simulations with RT using TRAPHIC, which is designed to handle large numbers of sources efficiently. We simulate the ionizing radiation from stars together with the UVB and recombination radiation. We show that the local stellar radiation can significantly change the H I column density distribution at column densities relevant for the so-called Damped Ly $\alpha$  (DLA) and Lyman-Limit (LL) systems. We also show that the main source of disagreement between previous works is insufficient resolution, a problem we solve by using star-forming particles as ionizing sources. We also show that the absence of a fully resolved ISM in cosmological simulations is a bottle-neck for modelling the properties of strong DLAs (i.e.,  $N_{\text{HI}} \geq 10^{21} \text{ cm}^{-2}$ ).

Strong H I absorbers, such as DLAs, are likely to be representative of the cold gas in, or close to, the ISM in high-redshift galaxies. Because of this, they provide a unique opportunity to define an absorption-selected galaxy sample and to study the ISM, particularly at the early stages of galaxy formation. However, because observational studies are limited by the small number of known strong H I absorbers and are missing low-mass galaxies in surveys for their counterparts in emission, it is very difficult to probe the relation between H I absorbers and their host galaxies observationally and we have to resort to cosmological simulations to help us understand the link between the two. In **Chapter 4**, we use the hydrodynamical cosmological simulations that we have shown to match the H I observations very well (see Chapter 2), to study the link between strong H I systems and galaxies at  $z = 3$ . We show that most strong H I absorbers are associated with low-mass galaxies too faint to be detected in current observations. We demonstrate, however, that our predictions are in good agreement with the existing observations. We show that there is a strong anti-correlation between the column density of strong H I absorbers and the impact parameters that connect them to their closest galaxies. We also investigate correlations between the column density of strong H I absorbers and different properties of their associated galaxies.

Similar to the neutral hydrogen, which provides fuel for star formation, the dust content of galaxies has a strong connection with their star formation activity. This makes studying the distribution and evolution of the dust also very important for understanding the evolution of galaxies. The low angular resolution of observations at long wavelengths makes identification and spectroscopy

of individual distant infrared galaxies a daunting task. Significant information about the evolution and statistical properties of these objects is encoded in the surface density of sources as a function of brightness (i.e., the source count). In **Chapter 5**, we present a model for the evolution of dusty galaxies, constrained by the  $850\mu\text{m}$  source counts and redshift distribution. We use a simple formalism for the evolution of the luminosity function and the color distribution of infrared galaxies. Using a novel algorithm for calculating the source counts, we analyze how individual free parameters in the model are constrained by observational data. The model is shown to successfully reproduce the observed source count and redshift distributions at wavelengths  $70\mu\text{m} \lesssim \lambda \lesssim 1100\mu\text{m}$ , and to be in excellent agreement with the most recent *Herschel* and *SCUBA 2* results.

## References

- al-Sufi, Abd al-Rahman, *Book of Fixed Stars*, 964, Isfahan, Iran
- Bagla, J. S. 2002, *Journal of Astrophysics and Astronomy*, 23, 185
- Becker, G. D., & Bolton, J. S. 2013, arXiv:1307.2259
- Bennett, C. L., Halpern, M., Hinshaw, G., et al. 2003, *ApJS*, 148, 1
- Gingold, R. A., & Monaghan, J. J. 1977, *MNRAS*, 181, 375
- Heymans, C., Grocutt, E., Heavens, A., et al. 2013, *MNRAS*, 432, 2433
- Kennicutt, R. C., Jr. 1998, *ApJ*, 498, 541
- Kim, T.-S., Carswell, R. F., Cristiani, S., D’Odorico, S., & Giallongo, E. 2002, *MNRAS*, 335, 555
- Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, *ApJS*, 192, 18
- Lucy, L. B. 1977, *AJ*, 82, 1013
- Mather, J. C., Cheng, E. S., Eplee, R. E., Jr., et al. 1990, *ApJL*, 354, L37
- Noterdaeme, P., Petitjean, P., Ledoux, C., & Srianand, R. 2009, *A&A*, 505, 1087
- O’Meara, J. M., Prochaska, J. X., Burles, S., et al. 2007, *ApJ*, 656, 666
- O’Meara, J. M., Prochaska, J. X., Worseck, G., Chen, H.-W., & Madau, P. 2013, *ApJ*, 765, 137
- Pawlik, A. H., & Schaye, J. 2008, *MNRAS*, 389, 651
- Pawlik, A. H., & Schaye, J. 2011, *MNRAS*, 412, 1943
- Péroux, C., Dessauges-Zavadsky, M., D’Odorico, S., Sun Kim, T., & McMahon, R. G. 2005, *MNRAS*, 363, 479
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2013, arXiv:1303.5076
- Prochaska, J. X., Worseck, G., & O’Meara, J. M. 2009, *ApJL*, 705, L113
- Schaye, J., & Dalla Vecchia, C. 2008, *MNRAS*, 383, 1210
- Schaye, J., Dalla Vecchia, C., Booth, C. M., et al. 2010, *MNRAS*, 402, 1536
- Spergel, D. N., Verde, L., Peiris, H. V., et al. 2003, *ApJS*, 148, 175
- Tytler, D. 1987, *ApJ*, 321, 49
- Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, *ApJ*, 736, 59



# 2

## ON THE EVOLUTION OF THE HI COLUMN DENSITY DISTRIBUTION IN COSMOLOGICAL SIMULATIONS

We use a set of cosmological simulations combined with radiative transfer calculations to investigate the distribution of neutral hydrogen in the post-reionization Universe. We assess the contributions from the metagalactic ionizing background, collisional ionization and diffuse recombination radiation to the total ionization rate at redshifts  $z = 0 - 5$ . We find that the densities above which hydrogen self-shielding becomes important are consistent with analytic calculations and previous work. However, because of diffuse recombination radiation, whose intensity peaks at the same density, the transition between highly ionized and self-shielded regions is smoother than what is usually assumed. We provide fitting functions to the simulated photoionization rate as a function of density and show that post-processing simulations with the fitted rates yields results that are in excellent agreement with the original radiative transfer calculations. The predicted neutral hydrogen column density distributions agree very well with the observations. In particular, the simulations reproduce the remarkable lack of evolution in the column density distribution of Lyman limit and weak damped Ly $\alpha$  systems below  $z = 3$ . The evolution of the low column density end is affected by the increasing importance of collisional ionization with decreasing redshift. On the other hand, the simulations predict the abundance of strong damped Ly $\alpha$  systems to broadly track the cosmic star formation rate density.

Alireza Rahmati, Andreas H. Pawlik, Milan Raičević, Joop Schaye  
*Monthly Notices of the Royal Astronomical Society*  
Volume 430, Issue 3, pp. 2427-2445 (2013)

## 2.1 Introduction

A substantial fraction of the interstellar medium (ISM) in galaxies consists of atomic hydrogen. This makes studying the distribution of neutral hydrogen (HI) and its evolution crucial for our understanding of various aspects of star formation. In the local universe, the HI content of galaxies is measured through 21-cm observations, but at higher redshifts this will not be possible until the advent of significantly more powerful telescopes such as the Square Kilometer Array<sup>1</sup>. However, at  $z \lesssim 6$ , i.e., after reionization, the neutral gas can already be probed through the absorption signatures imprinted by the intervening HI systems on the spectra of bright background sources, such as quasars (QSOs).

The early observational constraints on the HI column density distribution function (HI CDDF hereafter), from quasar absorption spectroscopy at  $z \lesssim 3$ , were well described by a single power-law in the range  $N_{\text{HI}} \sim 10^{13} - 10^{21} \text{ cm}^{-2}$  (Tytler, 1987). Thanks to a significant increase in the number of observed quasars and improved observational techniques, more recent studies have extended these observations to both lower and higher HI column densities and to higher redshifts (e.g., Kim et al., 2002; Péroux et al., 2005; O’Meara et al., 2007; Noterdaeme et al., 2009; Prochaska et al., 2009; Prochaska & Wolfe, 2009; O’Meara et al., 2013; Noterdaeme et al., 2012). These studies have revealed a much more complex shape which has been described using several different power-law functions (e.g., Prochaska et al., 2010; O’Meara et al., 2013).

The shape of the HI CDDF is determined by both the distribution and ionization state of hydrogen. Consequently, determining the distribution function of HI column densities requires not only accurate modeling of the cosmological distribution of gas, but also radiative transfer (RT) of ionizing photons. As a starting point, the HI CDDF can be modeled by assuming a certain gas profile and exposing it to an ambient ionizing radiation field (e.g., Petitjean et al., 1992; Zheng & Miralda-Escudé, 2002). Although this approach captures the effect of self-shielding, it cannot be used to calculate the detailed shape and normalization of the HI CDDF which results from the cumulative effect of large numbers of objects with different profiles, total gas contents, temperatures and sizes. Moreover, the interaction between galaxies and the circum-galactic medium through accretion and various feedback mechanisms, and its impact on the overall gas distribution are not easily captured by simplified models. Therefore, it is important to complement these models with cosmological simulations that model the evolution of the large-scale structure of the Universe and the formation of galaxies.

The complexity of the RT calculation depends on the HI column density. At low HI column densities (i.e.,  $N_{\text{HI}} \lesssim 10^{17} \text{ cm}^{-2}$ , corresponding to the so-called Lyman- $\alpha$  forest), hydrogen is highly ionized by the metagalactic ultraviolet background radiation (hereafter UVB) and largely transparent to the ion-

---

<sup>1</sup><http://www.skatelescope.org/>



izing radiation. For these systems, the HI column densities can therefore be accurately computed in the optically thin limit. At higher HI column densities (i.e.,  $N_{\text{HI}} \gtrsim 10^{17} \text{ cm}^{-2}$ , corresponding to the so-called Lyman Limit and Damped Lyman- $\alpha$  systems), the gas becomes optically thick and self-shielded. As a result, the accurate computation of the HI column densities in these systems requires precise RT simulations. On the other hand, at the highest HI column densities where the gas is fully self-shielded and the recombination rate is high, non-local RT effects are not very important and the gas remains largely neutral. At these column densities, the hydrogen ionization rate may, however, be strongly affected by the local sources of ionization (Miralda-Escudé, 2005; Schaye, 2006; Rahmati et al., 2013). In addition, other processes like H<sub>2</sub> formation (Schaye, 2001b; Krumholz et al., 2009; Altay et al., 2011) or mechanical feedback from young stars and / or AGNs (Erkal et al., 2012), can also affect the highest HI column densities.

Despite the importance of RT effects, most of the previous theoretical works on the HI column density distribution did not attempt to model RT effects in detail (e.g., Katz et al., 1996; Gardner et al., 1997; Haehnelt et al., 1998; Gardner et al., 2001; Cen et al., 2003; Nagamine et al., 2004, 2007). Only very recent works incorporated RT, primarily to account for the attenuation of the UVB (Razoumov et al., 2006; Pontzen et al., 2008; Fumagalli et al., 2011; Altay et al., 2011; McQuinn et al., 2011) and found a sharp transition between optically thin and self-shielded gas that is expected from the exponential nature of extinction.

The aforementioned studies focused mainly on redshifts  $z = 2 - 3$ , for which observational constraints are strongest, without investigating the evolution of the HI distribution. They found that the HI CDDF in current cosmological simulations is in reasonable agreement with observations in a large range of HI column densities. Only at the highest HI column densities (i.e.,  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ ) the agreement is poor. However, it is worth noting that the interpretation of these HI systems is complicated due to the complex physics of the ISM and ionization by local sources. Moreover, the observational uncertainties are also larger for these rare high  $N_{\text{HI}}$  systems.

In this chapter, we investigate the cosmological HI distribution and its evolution during the last  $\gtrsim 12$  billion years (i.e.,  $z \lesssim 5$ ). For this purpose, we use a set of cosmological simulations which include star formation, feedback and metal-line cooling in the presence of the UVB. These simulations are based on the Overwhelmingly Large Simulations (OWLS) presented in Schaye et al. (2010). To obtain the HI CDDF, we post-processed the simulations with RT, accounting for both ionizing UVB radiation and ionizing recombination radiation (RR). In contrast to previous works, we account for the impact of recombination radiation explicitly, by propagating RR photons. Using these simulations we study the evolution of the HI CDDF in the range of redshifts  $z = 0 - 5$  for column densities  $N_{\text{HI}} \gtrsim 10^{16} \text{ cm}^{-2}$ . We discuss how the individual contributions from the UVB, RR and collisional ionization to the total ionization rate shape the HI CDDF and assess their relative importance at different redshifts.

The structure of this chapter is as follows. In §2.2 we describe the details of the hydrodynamical simulations and of the RT, including the treatment of the UVB and recombination radiation. In §2.3 we present the simulated H I CDDF and its evolution and compare it with observations. In the same section we also discuss the contributions of different ionizing processes to the total ionization rate and provide fitting functions for the total photoionization rate as a function of density which reproduce the RT results. Finally, we conclude in §2.4.

## 2.2 Simulation techniques

### 2.2.1 Hydrodynamical simulations

We use density fields from a set of cosmological simulations performed using a modified version of the smoothed particle hydrodynamics code GADGET-3 (last described in Springel, 2005). The subgrid physics is identical to that used in the reference simulation of the OWLS project (Schaye et al., 2010). Star formation is pressure dependent and reproduces the observed Kennicutt-Schmidt law (Schaye & Dalla Vecchia, 2008). Chemical evolution is followed using the model of Wiersma et al. (2009a), which traces the abundance evolution of eleven elements by following stellar evolution assuming a Chabrier (2003) initial mass function. Moreover, a radiative heating and cooling implementation based on Wiersma et al. (2009b) calculates cooling rates element-by-element (i.e., using the above mentioned 11 elements) in the presence of the uniform cosmic microwave background and the UVB model given by Haardt & Madau (2001). About 40 per cent of the available kinetic energy in type II SNe is injected in winds with initial velocity of  $600 \text{ km s}^{-1}$  and a mass loading parameter  $\eta = 2$  (Dalla Vecchia & Schaye, 2008). Our tests show that varying the implementation of the kinetic feedback only changes the H I CDDF in the highest column densities ( $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ ). However, the differences caused by these variations are smaller than the evolution in the H I CDDF and observational uncertainties (see Altay et al. in prep.).

We adopt fiducial cosmological parameters consistent with the most recent WMAP 7-year results:  $\Omega_{\text{m}} = 0.272$ ,  $\Omega_{\text{b}} = 0.0455$ ,  $\Omega_{\Lambda} = 0.728$ ,  $\sigma_8 = 0.81$ ,  $n_s = 0.967$  and  $h = 0.704$  (Komatsu et al., 2011). We also use cosmological simulations from the OWLS project which are performed with a cosmology consistent with WMAP 3-year values with  $\Omega_{\text{m}} = 0.238$ ,  $\Omega_{\text{b}} = 0.0418$ ,  $\Omega_{\Lambda} = 0.762$ ,  $\sigma_8 = 0.74$ ,  $n_s = 0.951$  and  $h = 0.73$ . We use those simulations to avoid expensive resimulation with a WMAP 7-year cosmology. Instead, we correct for the difference in the cosmological parameters as explained in Appendix B.

Our simulations have box sizes in the range  $L = 6.25 - 100 \text{ comoving } h^{-1} \text{ Mpc}$  and baryonic particle masses in the range  $1.7 \times 10^5 h^{-1} M_{\odot} - 8.7 \times 10^7 h^{-1} M_{\odot}$ . The suite of simulations allows us to study the dependence of our results on the box size and mass resolution. Characteristic parameters of the simulations are

summarized in Table 2.1.

## 2.2.2 Radiative transfer with TRAPHIC

The RT is performed using TRAPHIC (Pawlik & Schaye, 2008, 2011). TRAPHIC is an explicitly photon-conserving RT method designed to transport radiation directly on the irregular distribution of SPH particles using its full dynamic range. Moreover, by tracing photon packets inside a discrete number of cones, the computational cost of the RT becomes independent of the number of radiation sources. TRAPHIC is therefore particularly well-suited for RT calculation in cosmological density fields with a large dynamical range in densities and large numbers of sources. In the following we briefly describe how TRAPHIC works. More details, as well as various RT tests, can be found in Pawlik & Schaye (2008, 2011).

The photon transport in TRAPHIC proceeds in two steps: the isotropic emission of photon packets with a characteristic frequency  $\nu$  by source particles and their subsequent directed propagation on the irregular distribution of SPH particles. The spatial resolution of the RT is set by the number of neighbors for which we generally use the same number of SPH neighbors used for the underlying hydrodynamical simulations, i.e.,  $N_{\text{ngb}} = 48$ .

After source particles emit photon packets isotropically to their neighbors, the photon packets travel along their propagation directions to other neighboring SPH particles which are inside their transmission cones. Transmission cones are regular cones with opening solid angle  $4\pi/N_{\text{TC}}$  and are centered on the propagation direction. The parameter  $N_{\text{TC}}$  sets the angular resolution of the RT, and we adopt  $N_{\text{TC}} = 64$ . We demonstrate convergence of our results with the angular resolution in Appendix C. Note that the transmission cones are defined locally at the transmitting particle, and hence the angular resolution of the RT is independent of the distance from the source.

It can happen that transmission cones do not contain any neighboring SPH particles. In this case, additional particles (virtual particles, ViPs) are placed inside the transmission cones to accomplish the photon transport. The ViPs, which enable the particle-to-particle transport of photons along any direction independent of the spatially inhomogeneous distribution of the particles, do not affect the SPH simulation and are deleted after the photon packets have been transferred.

An important feature of the RT with TRAPHIC is the merging of photon packets which guarantees the independence of the computational cost from the number of sources. Different photon packets which are received by each SPH particle are binned based on their propagation directions in  $N_{\text{RC}}$  reception cones. Then, photon packets with identical frequencies that fall in the same reception cone are merged into a single photon packet with a new direction set by the weighted sum of the directions of the original photon packets. Consequently, each SPH particle holds at most  $N_{\text{RC}} \times N_\nu$  photon packets, where  $N_\nu$  is the number of

**Table 2.1:** List of cosmological simulations used in this work. All the simulations use model ingredients identical to the reference simulation of Schaye et al. (2010). From left to right the columns show: simulation identifier; comoving box size; number of dark matter particles (there are equally many baryonic particles); initial baryonic particle mass; dark matter particle mass; comoving (Plummer-equivalent) gravitational softening; maximum physical softening; final redshift; cosmology. The last column shows whether the simulation was post-processed with RT. In simulations without RT, the H $\alpha$  distribution is obtained by using a fit to the photoionization rates as a function of density measured from simulations with RT.

Simulation	$L$ ( $h^{-1}\text{Mpc}$ )	$N$	$m_b$ ( $h^{-1}M_\odot$ )	$m_{\text{dm}}$ ( $h^{-1}M_\odot$ )	$\epsilon_{\text{com}}$ ( $h^{-1}\text{kpc}$ )	$\epsilon_{\text{prop}}$ ( $h^{-1}\text{kpc}$ )	$z_{\text{end}}$	Cosmology	RT
<i>L006N256</i>	6.25	$256^3$	$1.7 \times 10^5$	$7.9 \times 10^5$	0.98	0.25	2	WMAP7	✓
<i>L006N128</i>	6.25	$128^3$	$1.4 \times 10^6$	$6.3 \times 10^6$	1.95	0.50	0	WMAP7	✓
<i>L012N256</i>	12.50	$256^3$	$1.4 \times 10^6$	$6.3 \times 10^6$	1.95	0.50	2	WMAP7	✓
<i>L025N512</i>	25.00	$512^3$	$1.4 \times 10^6$	$6.3 \times 10^6$	1.95	0.50	2	WMAP7	✗
<i>L006N128-W3</i>	6.25	$128^3$	$1.4 \times 10^6$	$6.3 \times 10^6$	1.95	0.50	2	WMAP3	✓
<i>L025N512-W3</i>	25.00	$512^3$	$1.4 \times 10^6$	$6.3 \times 10^6$	1.95	0.50	2	WMAP3	✗
<i>L025N128-W3</i>	25.00	$128^3$	$8.7 \times 10^7$	$4.1 \times 10^8$	7.81	2.00	0	WMAP3	✓
<i>L050N256-W3</i>	50.00	$256^3$	$8.7 \times 10^7$	$4.1 \times 10^8$	7.81	2.00	0	WMAP3	✓
<i>L050N512-W3</i>	50.00	$512^3$	$1.1 \times 10^7$	$5.1 \times 10^7$	3.91	1.00	0	WMAP3	✓
<i>L100N512-W3</i>	100.00	$512^3$	$8.7 \times 10^7$	$4.1 \times 10^8$	7.81	2.00	0	WMAP3	✗

frequency bins. We set  $N_{\text{RC}} = 8$  for which our tests yield converged results.

Photon packets are transported along their propagation direction until they reach the distance they are allowed to travel within the RT time step by the finite speed of light, i.e.,  $c\Delta t$ . Photon packets that cross the simulation box boundaries are assumed to be lost from the computational domain. We use a time step  $\Delta t = 1 \text{ Myr} \left( \frac{L_{\text{box}}}{6.25 h^{-1} \text{ Mpc}} \right) \left( \frac{4}{1+z} \right) \left( \frac{128}{N_{\text{SPH}}} \right)$ , where  $N_{\text{SPH}}$  is the number of SPH particles in each dimension. We verified that our results are insensitive to the exact value of the RT time step: values that are smaller or larger by a factor of two produce essentially identical results. This is mostly because we evolve the ionization balance on smaller subcycling steps, and because we iterate for the equilibrium solution, as we discuss below. At the end of each time step the ionization states of the particles are updated based on the number of absorbed ionizing photons.

The number of ionizing photons that are absorbed during the propagation of a photon packet from one particle to its neighbor is given by  $\delta\mathcal{N}_{\text{abs},\nu} = \delta\mathcal{N}_{\text{in},\nu}[1 - \exp(-\tau(\nu))]$  where  $\delta\mathcal{N}_{\text{in},\nu}$  and  $\tau(\nu)$  are, respectively, the initial number of ionizing photons in the photon packet with frequency  $\nu$  and the total optical depth of all the absorbing species. In this work we mainly consider hydrogen ionization, but in general the total optical depth is the sum  $\tau(\nu) = \sum_{\alpha} \tau_{\alpha}(\nu)$  of the optical depth of each absorbing species (i.e.,  $\alpha \in \{\text{HI}, \text{HeI}, \text{HeII}\}$ ). Assuming that neighboring SPH particles have similar densities, we approximate the optical depth of each species using  $\tau_{\alpha}(\nu) = \sigma_{\alpha}(\nu)n_{\alpha}d_{\text{abs}}$ , where  $n_{\alpha}$  is the number density of species,  $d_{\text{abs}}$  is the absorption distance between the SPH particle and its neighbor and  $\sigma_{\alpha}(\nu)$  is the absorption cross section (Verner et al., 1996). Note that ViPs are deleted after each transmission, and hence the photons they absorb need to be distributed among their SPH neighbors. However, in order to decrease the amount of smoothing associated with this redistribution of photons, ViPs are assigned only 5 (instead of 48) SPH neighbors. We demonstrate convergence of our results with the number of ViP neighbors in Appendix C.

At the end of each RT time step, every SPH particle has a total number of ionizing photons that have been absorbed by each species,  $\Delta\mathcal{N}_{\text{abs},\alpha}(\nu)$ . This number is used in order to calculate the photoionization rate of every species for that SPH particle. For instance, the hydrogen photoionization rate is given by:

$$\Gamma_{\text{HI}} = \frac{\sum_{\nu} \Delta\mathcal{N}_{\text{abs,HI}}(\nu)}{\eta_{\text{HI}}\mathcal{N}_{\text{H}}\Delta t}, \quad (2.1)$$

where  $\mathcal{N}_{\text{H}}$  is the total number of hydrogen atoms inside the SPH particle and  $\eta_{\text{HI}} \equiv n_{\text{HI}}/n_{\text{H}}$  is the hydrogen neutral fraction.

Once the photoionization rate is known, the evolution of the ionization states is calculated. For instance, the equation which governs the ionization state of hydrogen is

$$\frac{d\eta_{\text{HI}}}{dt} = \alpha_{\text{HII}}n_e(1 - \eta_{\text{HI}}) - \eta_{\text{HI}}(\Gamma_{\text{HI}} + \Gamma_{e,\text{H}}n_e), \quad (2.2)$$

where  $n_e$  is the free electron number density,  $\Gamma_{e,H}$  is the collisional ionization rate and  $\alpha_{\text{HII}}$  is the HII recombination rate. The differential equations which govern the ionization balance (e.g., equation 2.2) are solved using a subcycling time step,  $\delta t = \min(f\tau_{\text{eq}}, \Delta t)$  where  $\tau_{\text{eq}} \equiv \tau_{\text{ion}}\tau_{\text{rec}}/(\tau_{\text{ion}} + \tau_{\text{rec}})$ , and  $f$  is a dimensionless factor which controls the integration accuracy (we set it to  $10^{-3}$ ),  $\tau_{\text{rec}} \equiv 1/\sum_i n_e\alpha_i$  and  $\tau_{\text{ion}} \equiv 1/\sum_i(\Gamma_i + n_e\Gamma_{e,i})$ . The subcycling scheme allows the RT time step to be chosen independently of the photoionization and recombination time scales without compromising the accuracy of the ionization state calculations<sup>2</sup>.

We employ separate frequency bins to transport UVB and RR photons. Because the propagation directions of photons in different frequency bins are merged separately, this allows us to track the individual radiation components, i.e., UVB and RR, and to compute their contributions to the total photoionization rate. The implementation of the UVB and the recombination radiation is described in § 2.2.3 and § 2.2.4 below.

At the start of the RT, the hydrogen is assumed to be neutral. In addition, we use a common simplification (e.g., Faucher-Giguère et al., 2009; McQuinn & Switzer, 2010; Altay et al., 2011) by assuming a hydrogen mass fraction of unity, i.e., we ignore helium (only for the RT). To calculate recombination and collisional ionizations rates, we set, in post-processing, the temperatures of star-forming gas particles with densities  $n_{\text{H}} > 0.1 \text{ cm}^{-3}$  to  $T_{\text{ISM}} = 10^4 \text{ K}$ , which is typical of the observed warm-neutral phase of the ISM. This is needed because in our hydrodynamical simulations the star-forming gas particles follow a polytropic equation of state which defines their effective temperatures. These temperatures are only a measure of the imposed pressure and do not represent physical temperatures (see Schaye & Dalla Vecchia, 2008). To speed up convergence, the hydrogen at low densities (i.e.,  $n_{\text{H}} < 10^{-3} \text{ cm}^{-3}$ ) or high temperatures (i.e.,  $T > 10^5 \text{ K}$ ) is assumed to be in ionization equilibrium with the UVB and the collisional ionization rate (see Appendix A2). Typically, the neutral fraction of the box and the resulting H<sub>I</sub> CDDF do not evolve after 2-3 light-crossing times (the light-crossing time for the extended box with  $L_{\text{box}} = 6.25 \text{ comoving } h^{-1}\text{Mpc}$  is  $\approx 7.5 \text{ Myr}$  at  $z = 3$ ).

### 2.2.3 Ionizing background radiation

Although our hydrodynamical simulations are performed using periodic boundary conditions, we use absorbing boundary conditions for the RT. This is necessary because our box size is much smaller than the mean free path of ionizing photons. We simulate the ionizing background radiation as plane-parallel radi-

---

<sup>2</sup>Other considerations prevent the use of arbitrarily large RT time steps. The RT assumes that species fractions and hence opacities do not evolve within a RT time step. This approximation becomes increasingly inaccurate with increasing RT time steps. Note that in this work, we iterate for ionization equilibrium which help to render our results robust against changes in the RT time step, as our convergence studies confirm.

**Table 2.2:** Hydrogen photoionization rate, absorption cross-section, equivalent gray approximation frequency and the self-shielding density threshold (i.e., based on equation 2.13) for three UVB models: Haardt & Madau (2001) (HM01; used in this work), Haardt & Madau (2012) (HM12) and Faucher-Giguère et al. (2009) (FG09) at different redshifts. For the calculation of the photoionization rate and absorption cross-sections, only photons with energies below 54.4 eV are taken into account, effectively assuming that more energetic photons are absorbed by He.

Redshift	UVB	$\Gamma_{\text{UVB}} \text{ (s}^{-1}\text{)}$	$\bar{\sigma}_{\text{vHI}} \text{ (cm}^2\text{)}$	$\mathcal{E}_{\text{eq}} \text{ (eV)}$	$n_{\text{H,SSh}} \text{ (cm}^{-3}\text{)}$
$z = 0$	HM01	$8.34 \times 10^{-14}$	$3.27 \times 10^{-18}$	16.9	$1.1 \times 10^{-3}$
	HM12	$2.27 \times 10^{-14}$	$2.68 \times 10^{-18}$	18.1	$5.1 \times 10^{-4}$
	FG09	$3.99 \times 10^{-14}$	$2.59 \times 10^{-18}$	18.3	$7.7 \times 10^{-4}$
$z = 1$	HM01	$7.39 \times 10^{-13}$	$2.76 \times 10^{-18}$	17.9	$5.1 \times 10^{-3}$
	HM12	$3.42 \times 10^{-13}$	$2.62 \times 10^{-18}$	18.2	$3.3 \times 10^{-3}$
	FG09	$3.03 \times 10^{-13}$	$2.37 \times 10^{-18}$	18.8	$3.1 \times 10^{-3}$
$z = 2$	HM01	$1.50 \times 10^{-12}$	$2.55 \times 10^{-18}$	18.3	$8.7 \times 10^{-3}$
	HM12	$8.98 \times 10^{-13}$	$2.61 \times 10^{-18}$	18.2	$6.1 \times 10^{-3}$
	FG09	$6.00 \times 10^{-13}$	$2.27 \times 10^{-18}$	19.1	$5.1 \times 10^{-3}$
$z = 3$	HM01	$1.16 \times 10^{-12}$	$2.49 \times 10^{-18}$	18.5	$7.4 \times 10^{-3}$
	HM12	$8.74 \times 10^{-13}$	$2.61 \times 10^{-18}$	18.2	$6.0 \times 10^{-3}$
	FG09	$5.53 \times 10^{-13}$	$2.15 \times 10^{-18}$	19.5	$5.0 \times 10^{-3}$
$z = 4$	HM01	$7.92 \times 10^{-13}$	$2.45 \times 10^{-18}$	18.6	$5.8 \times 10^{-3}$
	HM12	$6.14 \times 10^{-13}$	$2.60 \times 10^{-18}$	18.3	$4.7 \times 10^{-3}$
	FG09	$4.31 \times 10^{-13}$	$2.02 \times 10^{-18}$	19.9	$4.4 \times 10^{-3}$
$z = 5$	HM01	$5.43 \times 10^{-13}$	$2.45 \times 10^{-18}$	18.6	$4.5 \times 10^{-3}$
	HM12	$4.57 \times 10^{-13}$	$2.58 \times 10^{-18}$	18.3	$3.9 \times 10^{-3}$
	FG09	$3.52 \times 10^{-13}$	$1.94 \times 10^{-18}$	20.1	$4.0 \times 10^{-3}$



ation entering the simulation box from its sides. At the beginning of each RT step, we generate a large number of photon packets,  $N_{\text{bg}}$ , on the nodes of a regular grid at each side of the simulation box and set their propagation directions perpendicular to the sides. The number of photon packets is chosen to obtain converged results. Furthermore, to avoid numerical artifacts close to the edges of the box, we use the periodicity of our simulations to extend the simulation box by the typical size of the region where we generate the background radiation (i.e., 2% of the box size from each side). These extended regions are excluded from the analysis, thereby removing the artifacts without losing any information contained in the original simulation box.

The photon content of each packet is normalized such that in the absence of any absorption (i.e., assuming the optically thin limit), the total photon density of the box corresponds to the desired uniform hydrogen photoionization rate. If we assume that all the photons with frequencies higher than  $\nu_{\text{HeII}}$  are absorbed by helium, then the hydrogen photoionization rate can be written as:

$$\begin{aligned}\Gamma_{\text{UVB}} &= \int_{\nu_{\text{HI}}}^{\nu_{\text{HeII}}} 4\pi \frac{J_\nu}{h\nu} \sigma_{\text{HI},^\circ} d\nu \\ &\equiv \frac{4\pi \bar{\sigma}_{\nu_{\text{HI}}}}{h} \int_{\nu_{\text{HI}}}^{\nu_{\text{HeII}}} \frac{J_\nu}{\nu} d\nu,\end{aligned}\quad (2.3)$$

where  $J_\nu$  is the radiation intensity (in units  $\text{erg cm}^{-2} \text{s}^{-1} \text{sr}^{-1} \text{Hz}^{-1}$ ),  $\nu_{\text{HI}}$  and  $\nu_{\text{HeII}}$  are respectively the frequency at the Lyman-limit and the frequency at the HeII ionization edge, and  $\sigma_{\text{HI},^\circ}$  is the neutral hydrogen absorption cross-section for ionizing photons. In the last equation we have defined the gray absorption cross-section,

$$\bar{\sigma}_{\nu_{\text{HI}}} \equiv \frac{\int_{\nu_{\text{HI}}}^{\nu_{\text{HeII}}} J_\nu / \nu \sigma_{\text{HI},^\circ} d\nu}{\int_{\nu_{\text{HI}}}^{\nu_{\text{HeII}}} J_\nu / \nu d\nu}.\quad (2.4)$$

The radiation intensity is related to the photon energy density,  $u_\nu$ ,

$$J_\nu = \frac{u_\nu c}{4\pi} = \frac{n_\nu h\nu c}{4\pi},\quad (2.5)$$

where  $n_\nu$  is the number density of photons inside the box. Combining Equations 2.3-2.5 yields

$$\Gamma_{\text{UVB}} = n_{\nu_{\text{HI}}} c \bar{\sigma}_{\nu_{\text{HI}}},\quad (2.6)$$

where  $n_{\nu_{\text{HI}}}$  is the number density of ionizing photons inside the box. The total number of ionizing photons in the box is therefore given by

$$n_{\nu_{\text{HI}}} L_{\text{box}}^3 = n_\gamma 6 N_{\text{bg}} \frac{L_{\text{box}}}{c \Delta t},\quad (2.7)$$

where  $n_\gamma$  is the number of ionizing photons carried by each photon packet. Now we can calculate the photon content of each packet that must be injected into the



box during each step in order to achieve the desired H I photoionization rate:

$$n_\gamma = \frac{\Gamma_{\text{UVB}} L_{\text{box}}^2 \Delta t}{6 \bar{\sigma}_{\text{vHI}} N_{\text{bg}}}, \quad (2.8)$$

We use the redshift-dependent UVB spectrum of Haardt & Madau (2001) to calculate  $\Gamma_{\text{UVB}}$  and  $\bar{\sigma}_{\text{vHI}}$ . The Haardt & Madau (2001) UVB model successfully reproduces the relative strengths of the observed metal absorption lines in the intergalactic medium (Aguirre et al., 2008) and has been used to calculate heating/cooling in our cosmological simulations<sup>3</sup>. We note however that using more recent models for the UVB is not expected to change our main results. One can show that varying the UVB photoionization rate by a factor of 3, only changes the H I CDDF by less than 0.2 dex for LLSs (e.g., Altay et al., 2011). As shown in Table 2.2, the differences between photoionization rates in different UVB models are smaller than a factor of 3, particularly at  $z > 1$ , where the photoionization by the UVB is not subdominant (see §2.3.5). The variations in the adopted UVB model is even less important for systems with higher H I column densities (i.e., DLAs) which remain highly neutral for reasonable UVB models (e.g., Haardt & Madau, 2012; Faucher-Giguère et al., 2009).

To reduce the computational cost, we treat the multi-frequency problem in the gray approximation. In other words, we transport the UVB radiation using a single frequency bin, inside which photons are absorbed using the gray cross-section  $\bar{\sigma}_{\text{vHI}}$  defined in equation 2.4. Note that the gray approximation ignores the spectral hardening of the radiation field that would occur in multifrequency simulations. In Appendix D we show the result of repeating our simulations using multiple frequency bins, and also explicitly accounting for the absorption of photons by helium. These results clearly show the expected spectral hardening. The impact of spectral hardening on the hydrogen neutral fractions and the H I CDDF is small. However, we note that spectral hardening can change the temperature of the gas in self-shielded regions and that this effect is not captured in our simulations.

Hydrogen photoionization rates and average absorption cross-sections for UVB radiation at different redshifts are listed in Table 2.2 for our fiducial UVB model based on Haardt & Madau (2001) together with Haardt & Madau (2012) and Faucher-Giguère et al. (2009). The photoionization rate peaks at  $z \approx 2 - 3$  in those models and the equivalent effective photon energy<sup>4</sup> of the background radiation changes only weakly with redshift, compared to the total photoionization rate.

<sup>3</sup>Note that during the hydrodynamical simulations, photoheating from the UVB is applied to all gas particles. This ignores the self-shielding of hydrogen atoms against the UVB that occurs at densities  $n_{\text{H}} \gtrsim 10^{-3} - 10^{-2} \text{ cm}^{-3}$ . This inconsistency, which could affect both collisional ionization rates and the small-scale structure of the absorbers, has been found to have no significant impact on the simulated H I CDDF (Pontzen et al., 2008; McQuinn & Switzer, 2010; Altay et al., 2011).

<sup>4</sup>We defined the equivalent effective photon energy,  $\mathcal{E}_{\text{eq}}$ , which corresponds to the absorption cross section,  $\bar{\sigma}_{\text{vHI}}$ , as:  $\mathcal{E}_{\text{eq}} \equiv 13.6 \text{ eV} (\bar{\sigma}_{\text{vHI}} / \sigma_0)^{-1/3}$  where  $\sigma_0 = 6.3 \times 10^{-18} \text{ cm}^2$ .

## 2.2.4 Recombination radiation

Photons produced by the recombination of positive ions and electrons can also ionize the gas. If the recombining gas is optically thin, recombination radiation can escape and its ionizing effects can be ignored (i.e., the so-called Case A). However, for regions in which the gas is optically thick, the proper approximation is to assume the ionizing recombination radiation is absorbed on the spot. In this case, the effective recombination rate can be approximated by excluding the transitions that produce ionizing photons (e.g., Osterbrock & Ferland, 2006). This scenario is usually called Case B. A possible way to take into account the effect of recombination radiation is to use Case A recombination at low densities and Case B recombination at high densities (e.g., Altay et al., 2011; McQuinn et al., 2011), but this will be inaccurate in the transition regime.

In this work we explicitly treat the ionizing photons emitted by recombining hydrogen atoms and follow their propagation through the simulation box. This is facilitated by the fact that the computational cost of RT with TRAPHIC is independent of the number of sources. This is particularly important noting that every SPH particle is potentially a source. The photon production rates of SPH particles depend on their recombination rates and the radiation is emitted isotropically once at the beginning of every RT time step (see Raicevic et al. in prep. for full details).

We do not take into account the redshifting of the recombination photons by peculiar velocities of the emitters, or the Hubble flow. Instead, we assume that all recombination photons are monochromatic with energy 13.6 eV. In reality, recombination photons cannot travel to large cosmological distances without being redshifted to frequencies below the Lyman edge. Therefore, neglecting the cosmological redshifting of RR will result in overestimation of its photoionization rate on large scales. However, because of the small size of our simulation box, the total photoionization rate that is produced by RR on these scales remains negligible compared to the UVB photoionization rate. Consequently, the neglect of RR redshifting is not expected to affect our results.

## 2.2.5 The H $\alpha$ column density distribution function

In order to compare the simulation results with observations, we compute the CDDF of neutral hydrogen,  $f(N_{\text{HI}}, z)$ , a quantity that is somewhat straightforward to measure in QSO absorption line studies and is defined as the number of absorbers per unit column density, per unit absorption length,  $dX$ :

$$f(N_{\text{HI}}, z) \equiv \frac{d^2n}{dN_{\text{HI}}dX} \equiv \frac{d^2n}{dN_{\text{HI}}dz} \frac{H(z)}{H_0} \frac{1}{(1+z)^2}. \quad (2.9)$$

We project the HI content of the simulation box along each axis onto a grid with  $5000^2$  or  $10000^2$  pixels (for the  $128^3 - 256^3$  and  $512^3$  simulations, respectively)<sup>5</sup>. This is done using the actual kernels of SPH particles and for each of the three axes. The projection may merge distinct systems along the line of sight. However, for the small box sizes and high column densities with  $N_{\text{HI}} > 10^{17} \text{ cm}^{-2}$ , which are the focus of this work, the chance of overlap between multiple absorbing systems in projection is negligible. Based on our numerical experiments, we expect that this projection effect starts to appear only at  $N_{\text{HI}} < 10^{16} \text{ cm}^{-2}$  if one uses a single slice for the projection of the entire *L50N512-W3* simulation box at  $z = 3$ . To make sure our results are insensitive to this effect, we use, depending on redshift, 25 or 50 slices for projecting the *L50N512-W3* simulation.

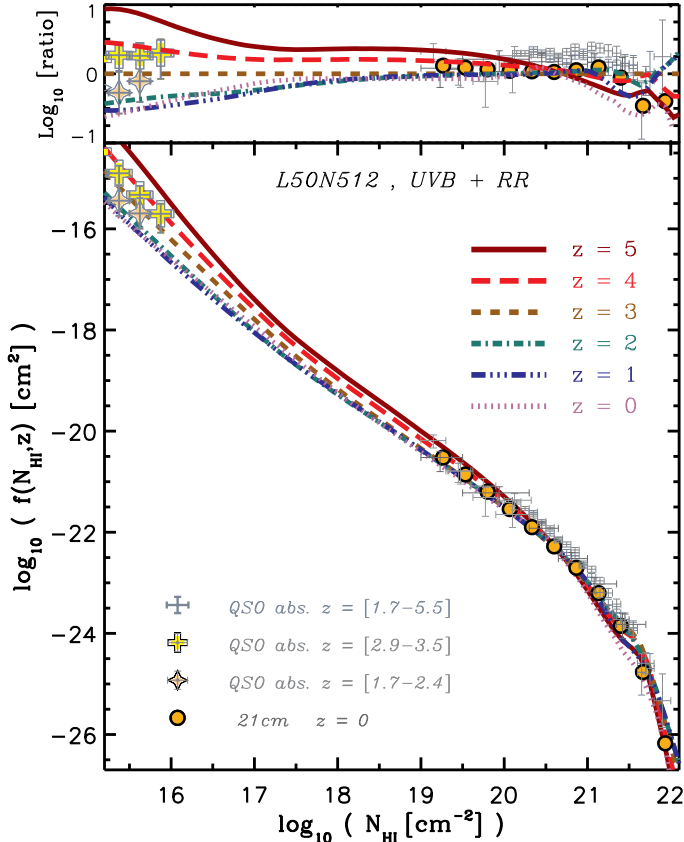
To produce a converged  $f(N_{\text{HI}}, z)$  from simulations, one needs to use cosmological boxes that are large enough to capture the relevant range of overdensities. This is particularly demanding at very high HI column densities: for instance, van de Voort et al. (2012) showed that most of the gas with  $N_{\text{HI}} > 10^{21} \text{ cm}^{-2}$  resides in galaxies with halo masses  $\gtrsim 10^{11} M_{\odot}$  which are relatively rare. As we show in Appendix B, the box size required to produce a converged HI CDDF up to  $N_{\text{HI}} \sim 10^{22} \text{ cm}^{-2}$  is  $L \gtrsim 50$  comoving  $h^{-1} \text{ Mpc}$ . Simulating RT in such a large volume is expensive. However, as we show in §2.3.4, at a given redshift the photoionization rates are fit very well by a function of the hydrogen number density. This relation is conserved with respect to both box size<sup>6</sup> and resolution and can therefore be applied to our highest resolution simulation (i.e., *L50N512-W3*), allowing us to keep the numerical cost tractable. Finally, since repeating the high-resolution simulations is expensive, we apply a redshift-independent correction which accounts for the difference between the WMAP year 3 parameters used for *L50N512-W3* simulation and the WMAP year 7 values. This is done by multiplying all the HI CDDFs produced based on the WMAP year 3 cosmology by the ratio between the HI CDDFs for *L25N512* and *L25N512-W3* at  $z = 3$ .

## 2.2.6 Dust and molecular hydrogen

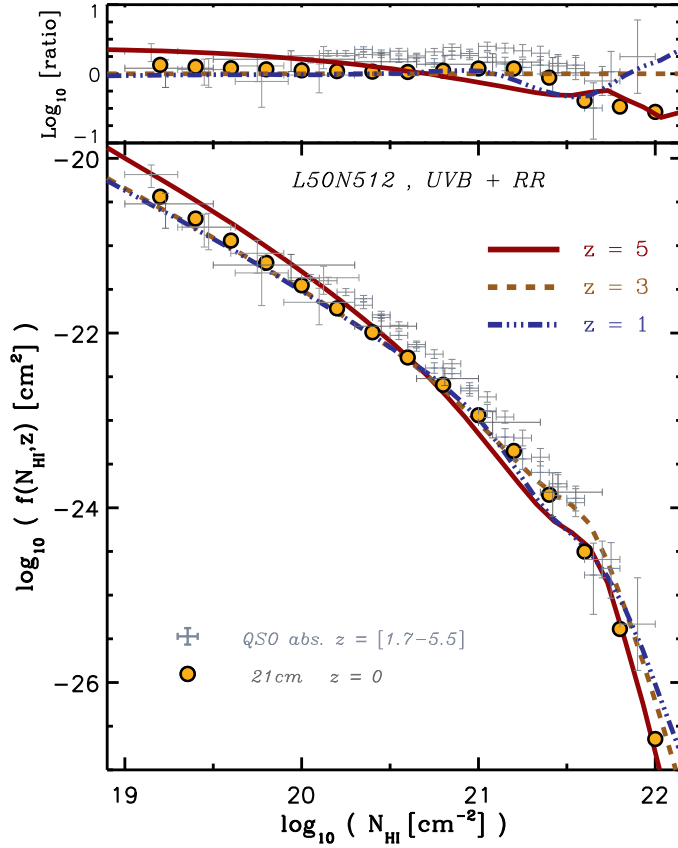
Dust and star formation are highly correlated and infrared observations indicate that the prevalence of dusty galaxies follows the average star formation history of the Universe (e.g., Rahmati & van der Werf, 2011). Nevertheless, dust extinction is a physical processes that is not treated in our simulations. Assuming a constant dust-to-gas ratio, the typical dust absorption cross-section per atom is orders of magnitudes lower than the typical hydrogen absorption cross-section

<sup>5</sup>Using  $5000^2$  cells, the corresponding cell size is similar to the minimum smoothing length, and  $\sim 100$  times smaller than the mean smoothing length, of SPH particles at  $z = 3$  in the *L06N128* simulation.

<sup>6</sup>One should note that the box size can indirectly change the resulting photoionization rate profile. For instance, self-shielding can be affected by collisional ionization, which become stronger at lower redshifts and whose importance depends on the abundance of massive objects, which is more sensitive to the box size.



**Figure 2.1:** CDDF of neutral gas at different redshifts in the presence of the UVB and diffuse recombination radiation for *L50N512-W3*. A column density dependent amplitude correction has been applied to make the results consistent with WMAP year 7 cosmological parameters. The observational data points represent a compilation of various quasar absorption line observations at high redshifts (i.e.,  $z = [1.7, 5.5]$ ) taken from Péroux et al. (2005) with  $z = [1.8, 3.5]$ , O’Meara et al. (2007) with  $z = [1.7, 4.5]$ , Noterdaeme et al. (2009) with  $z = [2.2, 5.5]$  and Prochaska & Wolfe (2009) with  $z = [2.2, 5.5]$ . The colored data points in the top-left corner of the left panel are taken from Kim et al. (2002) with  $z = [2.9, 3.5]$  and  $z = [1.7, 2.4]$  for the yellow crosses and orange diamonds, respectively. The orange filled circles show the best-fit based on the low-redshift 21-cm observations of Zwaan et al. (2005). The high column density end of the H $\text{I}$  distribution is magnified in the *right panel* and for clarity only the simulated H $\text{I}$  CDDF of redshifts  $z = 1, 3$  &  $5$  are shown. The *top-section* of each panel shows the ratio between the H $\text{I}$  CDDFs at different redshifts and the H $\text{I}$  CDDF at  $z = 3$ . The simulation results are in reasonably good agreement with the observations and, like the observations, show only a remarkably weak evolution for Lyman Limit and weak damped Ly $\alpha$  systems below  $z = 3$ .



**Figure 2.2:** The high column density end of the HI distribution shown in Figure 2.1 is magnified here. For clarity, only the simulated HI CDDF of redshifts  $z = 1, 3$  &  $5$  are shown. The *top-section* shows the ratio between the HI CDDFs at different redshifts and the HI CDDF at  $z = 3$ . The simulation results are in reasonably good agreement with the observations and, like the observations, show only a weak evolution for Lyman Limit and weak damped Ly $\alpha$  systems below  $z = 3$ .

for ionizing photons (Weingartner & Draine, 2001). In other words, the absorption of ionizing photons by dust particles is not significant compared to the absorption by the neutral hydrogen. Consequently, as also found in cosmological simulations with ionizing radiation (Gnedin et al., 2008), dust absorption does not noticeably alter the overall distribution of ionizing photons and hydrogen neutral fractions.

The observed cut-off in the abundance of very high  $N_{\text{HI}}$  systems may be related to the conversion of atomic hydrogen into  $\text{H}_2$  (e.g., Schaye, 2001b; Krumholz et al., 2009; Prochaska & Wolfe, 2009; Altay et al., 2011). Following Altay et al. (2011) and Duffy et al. (2012), we adopt an observationally driven scaling relation between gas pressure and hydrogen molecular fraction (Blitz & Rosolowsky, 2006) in post-processing, which reduces the amount of observable HI at high densities. This scaling relation is based on observations of low-redshift galaxies and may not cover the low metallicities relevant for higher redshifts. This could be an issue, since the HI- $\text{H}_2$  relation is known to be sensitive to the dust content and hence to the metallicity (e.g., Schaye, 2001b, 2004; Krumholz et al., 2009a).

## 2.3 Results

In this section we report our findings based on various RT simulations which include UVB ionizing radiation and diffuse recombination radiation from ionized gas. As we demonstrate in §2.3.3, the dependence of the photoionization rate on density obtained from our RT simulations shows a generic trend for different resolutions and box sizes. Therefore, we can use the results of RT calculations obtained from smaller boxes (e.g.,  $L06N128$  or  $L06N256$ ) which are computationally cheaper, to calculate the neutral hydrogen distribution in larger boxes. The last column of Table 2.1 indicates for which simulations this was done.

In the following, we will first present the predicted HI CDDF and compare it with observations. Next we discuss other aspects of our RT results and the effects of ionization by the UVB, recombination radiation and collisional ionization on the resulting HI distributions at different redshifts.

### 2.3.1 Comparison with observations

In Figure 2.1 & 2.2 we compare the simulation results with a compilation of observed HI CDDFs, after converting both to the WMAP year 7 cosmology. The data points with error bars show results from high-redshift ( $z = 1.7 - 5.5$ ) QSO absorption line studies and the orange filled circles show the fitting function reported by Zwaan et al. (2005) based on 21-cm observations of nearby galaxies. The latter observations only probe column densities  $N_{\text{HI}} \gtrsim 10^{19} \text{ cm}^{-2}$ .

We note that the OWLS simulations have already been shown to agree with observations by Altay et al. (2011), but only for  $z = 3$  and based on a different

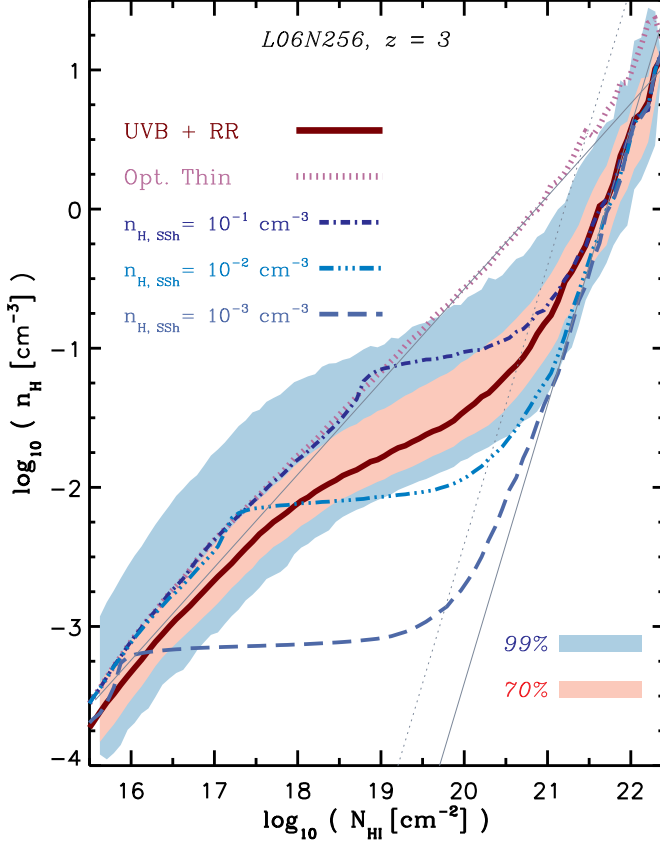
RT method (see Appendix C3 for a comparison). Overall, our RT results are also in good agreement with the observations. At high column densities (i.e.,  $N_{\text{HI}} > 10^{17} \text{ cm}^{-2}$ ) the observations probing  $0 < z < 5.5$  are consistent with each other. This implies weak or no evolution with redshift. The simulation is consistent with this remarkable observational result, predicting only weak evolution ( $\lesssim 50\%$ ) for  $10^{17} \text{ cm}^{-2} < N_{\text{HI}} < 10^{21} \text{ cm}^{-2}$  (i.e., Lyman limit systems, LLSs, and weak Damped Ly- $\alpha$  systems, DLAs) especially at  $z \lesssim 3$ .

The simulation predicts some variation with redshift for strong DLAs ( $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ ). The abundance of strong DLAs in the simulations follows a similar redshift-dependent trend as the average star formation density in our simulations which peaks at  $z \approx 2 - 3$  (Schaye et al., 2010). This result is consistent with the DLA evolution found by Cen (2012) in two zoomed simulations of a cluster and a void. One should, however, note that at very high column densities (e.g.,  $N_{\text{HI}} \gtrsim 10^{21.5} \text{ cm}^{-2}$ ) both observations and simulations are limited by small number statistics and the simulation results are more sensitive to the adopted feedback scheme (Altay et al. in prep.). Moreover, as we will show in Rahmati et al. (2013), including local stellar ionizing radiation can decrease the HI CDDF by up to  $\approx 1$  dex for  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ , especially at redshifts  $z \approx 2 - 3$  for which the average star formation activity of the Universe is near its peak.

At low column densities (i.e.,  $N_{\text{HI}} \lesssim 10^{17} \text{ cm}^{-2}$ ) the simulation results agree very well with the observations. This is apparent from the agreement between the simulated  $f(N_{\text{HI}}, z)$  at  $z = 3$  and  $z = 4$ , and the observed values for redshifts  $2.9 < z < 3.5$  (Kim et al., 2002) which are shown by the yellow crosses in Figure 2.1. The simulated  $f(N_{\text{HI}}, z)$  at lower and higher redshifts deviate from those at  $z \approx 3$  showing the abundance of those systems decreases with decreasing redshift and remains nearly constant at  $z \lesssim 2$ . This is consistent with the Ly- $\alpha$  forest observations at lower redshifts (Kim et al., 2002; Janknecht et al., 2006; Lehner et al., 2007; Prochaska & Wolfe, 2009; Ribaudo et al., 2011), as illustrated with the orange diamonds which correspond to  $z \approx 2$  observations, in the top-left corner of Figure 2.1.

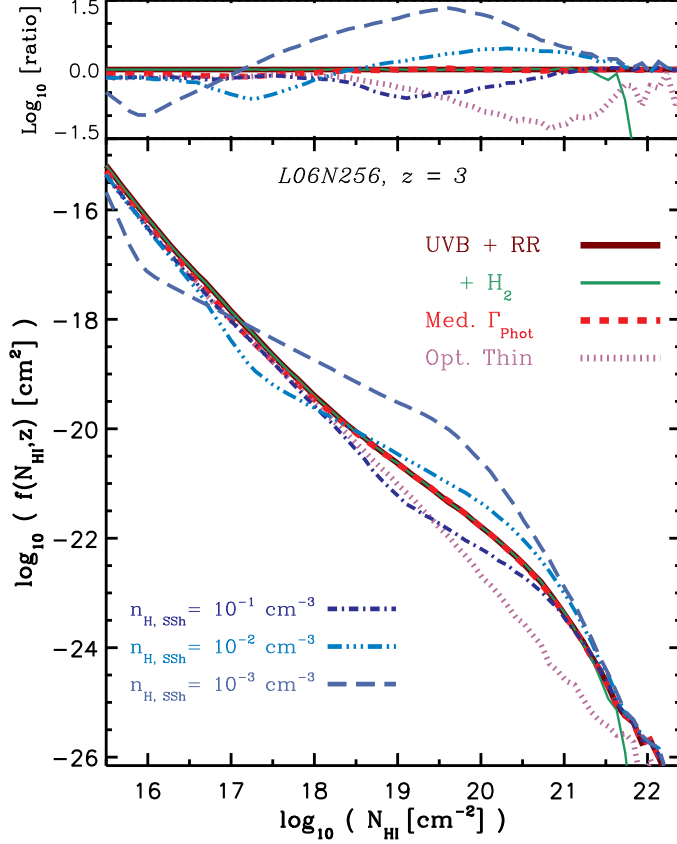
The evolution of the HI CDDF with redshift results from a combination of the expanding Universe and the growing intensity of the UVB radiation down to redshifts  $z \approx 2 - 3$ . At low redshifts (i.e.,  $z \approx 0$ ) the intensity of the UVB radiation has dropped by more than one order of magnitude leading to higher hydrogen neutral fractions and higher HI column densities. However, as we show in §2.3.5, at lower redshifts an increasing fraction of low-density gas is shock-heated to temperatures sufficiently high to become collisionally ionized and this compensates for the weaker UVB radiation at low redshifts.

The simulated HI CDDFs at all redshifts are consistent with each other and the observations. However, as illustrated in Figure 2.2, there is a  $\approx 0.2$  dex difference between the simulation results and the observations of LLS and DLAs at all redshifts. We found that the normalization of the HI CDDF in those regimes is sensitive to the adopted cosmological parameters (see Appendix B). Notably, the cosmology consistent with the WMAP 7 year results that is shown here, pro-



**Figure 2.3:**  $n_{\text{HI}}$ -weighted total hydrogen number density as a function of  $N_{\text{HI}}$ . The brown solid curve shows the RT results and the purple dotted curve shows the optically thin limit. Blue dot-dashed, dot-dot-dot-dashed and long dashed curves assume models with self-shielding density thresholds of  $n_{\text{H,SSh}} = 10^{-1}$ ,  $10^{-2}$  and  $10^{-3} \text{ cm}^{-3}$ , respectively. All of the above mentioned curves show the median of the  $n_{\text{HI}}$ -weighted total hydrogen number density at a given  $N_{\text{HI}}$ . The gray thin lines show the expected Jeans scaling relations for optically thin gas (equation 2.10; diagonal solid line) and for neutral gas (equation 2.11; steeper dotted line). A second solid line with the same slope expected from equation (2.11) but a different normalization is illustrated by the second solid line which is identical to the dotted line but shifted by 0.5 dex to higher  $N_{\text{HI}}$ . The pink and blue shaded areas in the right panel indicate the 70% and 99% scatter respectively, while the solid curves shows the median for the RT result. All the other curves are also medians. This shows that the  $n_{\text{H}} - N_{\text{HI}}$  relationship can be explained by the Jeans scaling and that the flattening in the CDDF is due to self-shielding.





**Figure 2.4:** HI CDDF in the presence of the UVB and diffuse recombination radiation for simulation *L06N256*. Simulations shown with different curves are identical to those in Figure 2.3. In addition, the effect of  $\text{H}_2$  formation is shown by the green solid curve which deviates from the brown solid curve at  $N_{\text{HI}} \gtrsim 3 \times 10^{21} \text{ cm}^{-2}$ . Finally the red dashed curve, which is indistinguishable from the brown solid curve, shows the result of assuming the median of the photoionization rate profile of the RT results to calculate the neutral fractions (see §2.3.3 and Appendix A1). The *top-section* in the right panel shows the ratio between different HI CDDFs and the one resulting from the RT simulations.

duces a better match to the observations than a cosmology based on the WMAP 3 year results with smaller values for  $\Omega_b$  and  $\sigma_8$ . This suggests that a higher value of  $\sigma_8$  may explain the small discrepancy between the simulation results and the observations.

### 2.3.2 The shape of the H I CDDF

The shape of the H I CDDF is determined by the distribution of hydrogen and by the different ionizing processes that set the hydrogen neutral fractions of the absorbers. One can assume that over-dense hydrogen resides in self-gravitating systems that are in local hydrostatic equilibrium. Then, the typical scales of the systems can be calculated as a function of the gas density based on a Jeans scaling argument (Schaye, 2001). Assuming that absorbers have universal baryon fractions (i.e.,  $f_g = \frac{\Omega_b}{\Omega_m}$ ) and typical temperatures of  $T_4 \equiv (T/10^4 \text{ K}) \sim 1$  (i.e., collisional ionization is unimportant), one can calculate the total hydrogen column density (Schaye, 2001):

$$N_{\text{H}} \sim 1.6 \times 10^{21} \text{ cm}^{-2} n_{\text{H}}^{1/2} T_4^{1/2} \left( \frac{f_g}{0.17} \right)^{1/2}. \quad (2.10)$$

Assuming that the gas is highly ionized and in ionization equilibrium with the ambient ionizing radiation field with the photoionization rate,  $\Gamma_{-12} = \Gamma/10^{-12} \text{ s}^{-1}$ , one gets (Schaye, 2001):

$$N_{\text{HI}} \sim 2.3 \times 10^{13} \text{ cm}^{-2} \left( \frac{n_{\text{H}}}{10^{-5} \text{ cm}^{-3}} \right)^{3/2} \times T_4^{-0.26} \Gamma_{-12}^{-1} \left( \frac{f_g}{0.17} \right)^{1/2}. \quad (2.11)$$

At high densities where the gas is nearly neutral, equation (2.10) provides a relation between  $N_{\text{HI}}$  and  $n_{\text{H}}$ . Equation (2.11) on the other hand, gives the relation for optically thin, highly ionized gas. The latter is derived assuming that the UVB photoionization is the dominant source of ionization, which is a good assumption at high redshifts and explains the relation between density and column density in Ly $\alpha$  forest simulations (e.g., Davé et al., 2010; Altay et al., 2011; McQuinn et al., 2011; Tepper-García et al., 2012). However, as we will show in the following sections, photoionization domination breaks down at lower redshifts where collisional ionization plays a significant role.

The column density at which hydrogen starts to be self-shielded against the UVB radiation follows from setting  $\tau_{\text{HI}} = 1$ :

$$N_{\text{HI,SSH}} \sim 4 \times 10^{17} \text{ cm}^{-2} \left( \frac{\bar{\sigma}_{v_{\text{HI}}}}{2.49 \times 10^{-18} \text{ cm}^2} \right)^{-1} \quad (2.12)$$

which can be used together with equation (2.11) to find the typical densities at

which the self-shielding begins (e.g., Furlanetto et al., 2005):

$$n_{\text{H,SSh}} \sim 6.73 \times 10^{-3} \text{ cm}^{-3} \left( \frac{\bar{\sigma}_{\nu_{\text{HI}}}}{2.49 \times 10^{-18} \text{ cm}^2} \right)^{-2/3} \times T_4^{0.17} \Gamma_{-12}^{2/3} \left( \frac{f_{\text{g}}}{0.17} \right)^{-1/3}. \quad (2.13)$$

These relations are compared with the  $n_{\text{H}}$ -weighted total hydrogen number density as a function of  $N_{\text{HI}}$  in the *L06N256* simulation at  $z = 3$  in Figure 2.3. The solid curve shows the median and the red (blue) shaded area represents the central 70% (90%) percentile. The diagonal gray solid line which converges with the simulation results at low column densities, shows equation (2.11) and the steeper gray dotted line which converges with the simulation results at high column densities is based on equation (2.10). The agreement between the expected slopes of the  $n_{\text{H}} - N_{\text{HI}}$  relation and the simulations at low and high column densities confirms our initial assumption that hydrogen resides in self-gravitating systems which are close to local hydrostatic equilibrium<sup>7</sup>.

As expected from equation (2.13), at low densities the gas is optically thin and follows the Jeans scaling relation of the highly ionized gas. At  $n_{\text{H}} \gtrsim 0.01 \text{ cm}^{-3}$  however, the relation between density and column density starts to deviate from equation (2.11) and approaches that of a nearly neutral gas. Consequently, for densities above the self-shielding threshold the HI column density increases rapidly over a narrow range of densities, leading to a flattening in the  $n_{\text{H}} - N_{\text{HI}}$  relation and in the resulting HI CDDF at  $N_{\text{HI}} \gtrsim 10^{18} \text{ cm}^{-2}$  (see Figure 2.4). The results from the RT simulation deviate from the magenta dotted lines, which are obtained assuming optically thin gas, at  $N_{\text{HI}} \gtrsim 4 \times 10^{17} \text{ cm}^{-2}$ . As the dotted line in Figure 2.4 shows, in the absence of self-shielding, the slope of  $f(N_{\text{HI}}, z) \propto N_{\text{HI}}^{\beta}$  is constant all the way up to DLAs at  $\beta_{\text{Ly}\alpha} \approx -1.6$ . However, because of self-shielding, the HI CDDF flattens to  $\beta_{\text{LLS}} \approx -1.1$  at  $10^{18} \text{ cm}^{-2} \lesssim N_{\text{HI}} \lesssim 10^{20} \text{ cm}^{-2}$  in the RT simulation (solid curve). These predicted slopes are in excellent agreement with the latest observational constraints of  $\beta_{\text{Ly}\alpha} \gtrsim -1.6$  for  $10^{15} \text{ cm}^{-2} < N_{\text{HI}} < 10^{17} \text{ cm}^{-2}$  to  $\beta_{\text{LLS}} \approx -1$  in the LLS regime (O’Meara et al., 2013). We also note that  $\beta_{\text{Ly}\alpha} \gtrsim -1.6$  is predicted to be almost the same for all redshifts, which agrees well with observations (Janknecht et al., 2006; Lehner et al., 2007; Ribaudó et al., 2011).

At densities  $n_{\text{H}} \gtrsim 0.1 \text{ cm}^{-3}$  the gas is nearly neutral and the Jeans scaling in equation (2.10) controls the  $n_{\text{H}} - N_{\text{HI}}$  relation. Consequently, the rate at which  $N_{\text{HI}}$  responds to changes in  $n_{\text{H}}$  slows down, causing a steepening in the resulting  $f(N_{\text{HI}}, z)$  in the DLA range (i.e.,  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ ). However, as the thick solid

<sup>7</sup>One should note that the above mentioned Jeans argument provides an order of magnitude calculation due to its simplifying assumptions (e.g., uniform density, universal baryon fraction, etc.). Although we may expect the predicted scaling relations to be correct, the very close agreement of the normalization with the simulations at low densities is coincidental. As the steeper gray dotted line which is based on equation (2.10) shows, the simulated  $N_{\text{HI}}$  for a given  $n_{\text{H}}$  is  $\approx 0.5$  dex higher than implied by the Jeans scaling for the nearly neutral case (i.e., steep, gray solid line).

curve in Figure 2.4 illustrates, the slope of  $f(N_{\text{HI}}, z)$  remains constant for  $N_{\text{HI}} = 10^{21} - 10^{22} \text{ cm}^{-2}$ . This is in contrast with observed trends indicating a sharp cut-off at  $N_{\text{HI}} \gtrsim 3 \times 10^{21} \text{ cm}^{-2}$  (Prochaska et al., 2010; O’Meara et al., 2013; but see Noterdaeme et al., 2012). At those column densities a large fraction of hydrogen is expected to form  $\text{H}_2$  molecules and be absent from HI observations (Schaye, 2001b; Krumholz et al., 2009; Altay et al., 2011). As the thin solid line in Figure 2.4 shows, accounting for  $\text{H}_2$  using the empirical relation between  $\text{H}_2$  fraction and pressure, based on  $z = 0$  observations (Blitz & Rosolowsky, 2006), does reproduce a sharp cut-off. If the observed relation does not cut off sharply (Noterdaeme et al., 2012), then this may imply that  $\text{H}_2$  fractions are lower at  $z = 3$  than at  $z = 0$ . We also note that the ionizing effect of local sources (Rahmati et al., 2013), increasing the efficiency of stellar feedback, e.g., by using a top-heavy IMF, and AGN feedback can also affect these high HI column densities (Altay et al. in prep.).

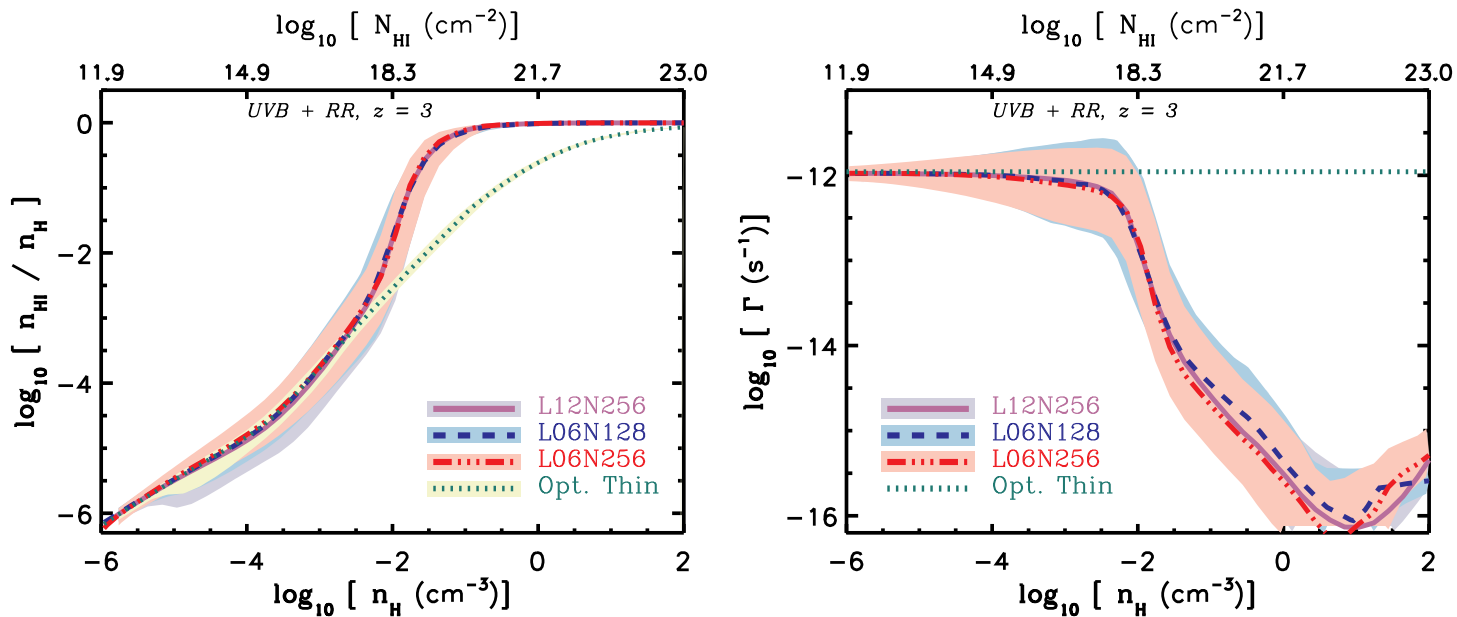
To first order, one can mimic the effect of RT by assuming gas with  $n_{\text{H}} < n_{\text{H,SSH}}$  to be optically thin (i.e., Case A recombination) and gas with  $n_{\text{H}} > n_{\text{H,SSH}}$  to be fully neutral. Simulations with three different self-shielding density thresholds are shown in Figure 2.3 & 2.4. The dot-dashed, dot-dot-dot-dashed and long dashed curves correspond to  $n_{\text{H,SSH}} = 10^{-1}$ ,  $10^{-2}$  and  $10^{-3} \text{ cm}^{-3}$ , respectively. Although all of these simulations predict the flattening of  $f(N_{\text{HI}}, z)$ , they produce a transition between optically thin and neutral gas that is too steep. In contrast, the RT results show a transition between highly ionized and highly neutral gas that is more gradual, as observed.

### 2.3.3 Photoionization rate as a function of density

Figure 2.5 illustrates the RT results for neutral fractions and photoionization rates as a function of density in the presence of UVB radiation and diffuse recombination radiation for the *L06N128*, *L06N256* and *L12N256* simulations at  $z = 3$ . For comparison, the results for the optically thin limit are shown by the green dotted curves. The sharp transition between highly ionized and neutral gas and its deviation from the optically thin case are evident in the left panel. This transition can also be seen in the photoionization rate (right panel) which drops at  $n_{\text{H}} \gtrsim 0.01 \text{ cm}^{-3}$ , consistent with equation (2.13) and previous studies (Tajiri & Umemura, 1998; Razoumov et al., 2006; Faucher-Giguère et al., 2010; Nagamine et al., 2010; Fumagalli et al., 2011; Altay et al., 2011).

The medians and the scatter around them are insensitive to the resolution of the underlying simulation and to the box size. This suggests that one can use the photoionization rate profile obtained from the RT simulations for calculating the hydrogen neutral fractions in other simulations for which no RT has been performed.

Moreover, as we show in §2.3.5, the total photoionization rate as a function of the hydrogen number density has the same shape at different redshifts. This shape can be characterized by three features: i) a knee at densities around the



**Figure 2.5:** The hydrogen neutral fraction (*left*) and the photoionization rate (*right*) as a function of hydrogen number density do not change by varying the simulation box size or mass resolution. This is shown for different simulations at  $z = 3$  in the presence of the UVB and recombination radiation. Purple solid, blue dashed and red dot-dashed lines show, respectively, the results for  $L12N256$ ,  $L06N128$  and  $L06N256$ . The green dotted line indicates the results for the  $L06N128$  simulation if the gas is assumed to be optically thin to the UVB radiation (i.e., no RT calculation is performed). The deviation between the optically thin hydrogen neutral fractions and RT results at  $n_{\text{H}} \geq 10^{-2} \text{ cm}^{-3}$  shows the impact of self-shielding. The lines show the medians and the shaded areas indicate the 15% – 85% percentiles. At the top of each panel we show HI column densities corresponding to each density.

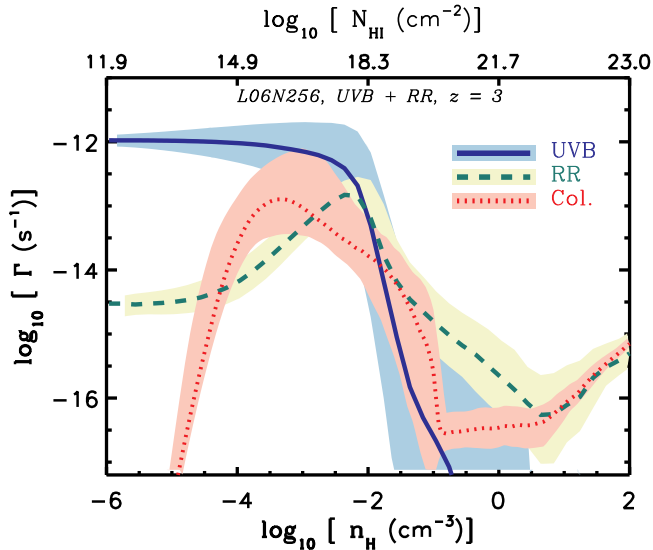
self-shielding density threshold, ii) a relatively steep fall-off at densities higher than the self-shielding threshold and iii) a flattening in the fall-off after the photoionization rate has dropped by  $\sim 2$  dex from its maximum value which is caused by the RR photoionization. These features are captured by the following fitting formula:

$$\frac{\Gamma_{\text{Phot}}}{\Gamma_{\text{UVB}}} = 0.98 \left[ 1 + \left( \frac{n_{\text{H}}}{n_{\text{H,SSh}}} \right)^{1.64} \right]^{-2.28} + 0.02 \left[ 1 + \frac{n_{\text{H}}}{n_{\text{H,SSh}}} \right]^{-0.84}, \quad (2.14)$$

where  $\Gamma_{\text{UVB}}$  is the background photoionization rate and  $\Gamma_{\text{Phot}}$  is the total photoionization rate. Moreover, the self-shielding density threshold,  $n_{\text{H,SSh}}$ , is given by equation (2.13) and is thus a function of  $\Gamma_{\text{UVB}}$  and  $\bar{\sigma}_{\text{vHI}}$  which vary with redshift. As explained in more detail in Appendix A1, the numerical parameters representing the shape of the profile are chosen to provide a redshift independent best fit to our RT results. In addition, the parametrization is based on the main RT related quantities, namely the intensity of UVB radiation and its spectral shape. It can therefore be used for UVB models similar to the Haardt & Madau (2001) model we used in this work (e.g., Faucher-Giguère et al., 2009; Haardt & Madau, 2012). For a given UVB model, one only needs to know  $\Gamma_{\text{UVB}}$  and  $\bar{\sigma}_{\text{vHI}}$  in order to determine the corresponding  $n_{\text{H,SSh}}$  from (2.13) (see also Table 2.2). Then, after using equation (2.14) to calculate the photoionization rate as a function of density, the equilibrium hydrogen neutral fraction for different densities, temperatures and redshifts can be readily calculated as explained in Appendix A1.

We note that the parameters used in equation (2.14) are only accurate for photoionization dominated cases. As we show in §2.3.5, at  $z \sim 0$  the collisional ionization rate is greater than the total photoionization rate around the self-shielding density threshold. Consequently, equation (2.13) does not provide an accurate estimate of the self-shielding density threshold at low redshifts. In Appendix A1 we therefore report the parameters that best reproduce our RT results at  $z = 0$ . Our tests show that simulations that use equation (2.14) reproduce the  $f(N_{\text{HI}}, z)$  accurately to within 10% for  $z \gtrsim 1$  where photoionization is dominant (see Appendix A1).

Although using the relation between the median photoionization rate and the gas density is a computationally efficient way of calculating equilibrium neutral fractions in big simulations, it comes at the expense of the information encoded in the scatter around the median photoionization rate at a given density. However, our experiments show that the error in  $f(N_{\text{HI}}, z)$  that results from neglecting the scatter in the photoionization rate profile is negligible for  $N_{\text{HI}} \gtrsim 10^{18} \text{ cm}^{-3}$  and less than  $\lesssim 0.1$  dex at lower column densities (see Appendix A1).

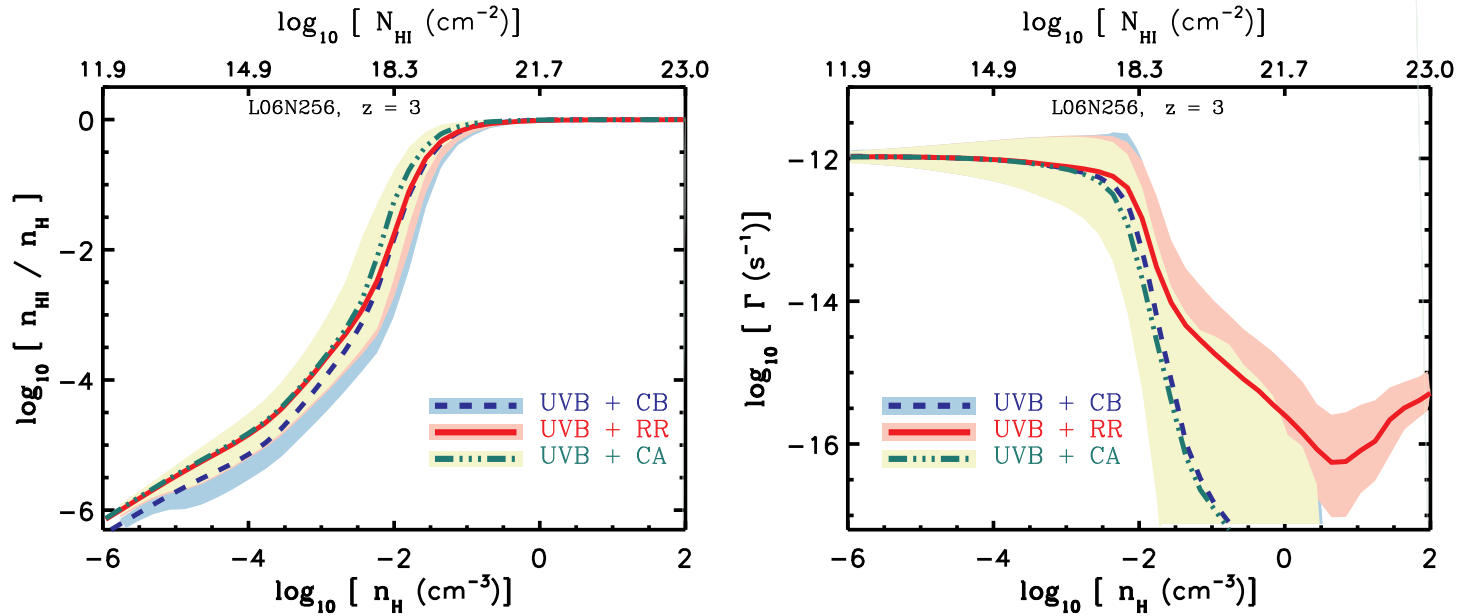


**Figure 2.6:** Ionization rates due to different sources of ionization as a function of hydrogen number density. Blue solid, green dashed and red dotted curves show, respectively, the UVB photoionization rate, the recombination radiation photoionization rate and the collisional ionization rate. The curves show the medians and the shaded areas around the medians indicate the 15% – 85% percentiles. HI column densities corresponding to each density are shown along the top x-axis. While the UVB is the dominant source of ionization below the self-shielding (i.e.,  $n_{\text{H}} \lesssim 10^{-2} \text{ cm}^{-3}$ ), recombination radiation dominates the ionization at higher densities.

### 2.3.4 The roles of diffuse recombination radiation and collisional ionization at $z = 3$

To study the interplay between different ionizing processes and their effects on the distribution of HI, we compare their ionization rates at different densities. We start the analysis by presenting the results at  $z = 3$  and extend it to other redshifts in §2.3.5.

The total photoionization rate profiles shown in the right panel of Figure 2.5 are almost flat at low densities and decrease with increasing density, starting at densities  $n_{\text{H}} \sim 10^{-4} \text{ cm}^{-3}$ . Just below  $n_{\text{H}} = 10^{-2} \text{ cm}^{-3}$  self-shielding causes a sharp drop, but the fall-off becomes shallower for  $n_{\text{H}} > 10^{-2} \text{ cm}^{-3}$  and the photoionization rate starts to increase at  $n_{\text{H}} > 10 \text{ cm}^{-3}$ . As shown in Figure 2.6, the shallower fall-off in the total photoionization rate with increasing density is caused by RR. The increase in the photoionization rate with density at the highest densities on the other hand, is an artifact of the imposed temperature for ISM particles (i.e.,  $T = 10^4 \text{ K}$ ) which produces a rising collisional ionization rate with increasing density. As the comparison between the UVB and RR photoionization profiles shows (see Figure 2.6), RR only starts to dominate the



**Figure 2.7:** The hydrogen neutral fractions (*left*) and the UVB photoionization rate profiles (*right*) as a function of density in RT simulations with different models for recombination radiation for the *L06N256* simulation at  $z = 3$ . The red dashed curve shows the reference simulation where recombination radiation is modeled self-consistently. The blue solid and green dot-dashed curves show simulations in which recombination radiation is substituted by the use of Case A and Case B recombination rates, respectively. The curves show the medians and the shaded areas around the medians indicate the 15% – 85% percentiles. H I column densities corresponding to each density are shown along the top x-axes. The effect of recombination radiation on the hydrogen neutral fractions is similar to the use of Case A recombination at low densities (i.e.,  $n_{\text{H}} \lesssim 10^{-3} \text{ cm}^{-3}$ ) and to the use of Case B recombination at higher densities (i.e.,  $n_{\text{H}} \gtrsim 10^{-1} \text{ cm}^{-3}$ ). However, recombination radiation can penetrate into the self-shielded regions, an effect that is not captured by the use of Case B recombination.



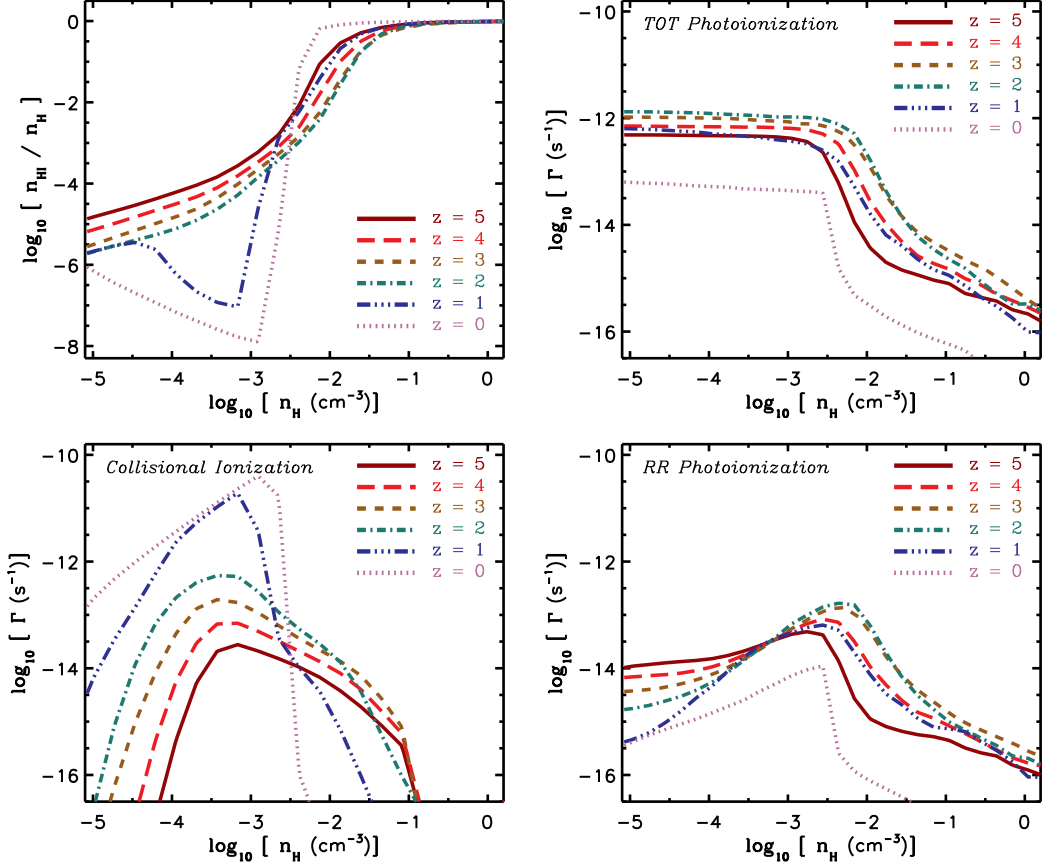
total photoionization rate at  $n_{\text{H}} > 10^{-2} \text{ cm}^{-3}$ , where the UVB photoionization rate has dropped by more than one order of magnitude and the gas is no longer highly ionized. RR reduces the total HI content of high-density gas by  $\approx 20\%$ . Although ionization rates remain non-negligible at higher densities, they cannot keep the hydrogen highly ionized. For instance at  $n_{\text{H}} \sim 1 \text{ cm}^{-3}$ , a photoionization rate of  $\Gamma \sim 10^{-14} \text{ s}^{-1}$  can only ionize the gas by  $\lesssim 20\%$ .

The shape of the photoionization rate profile produced by diffuse RR can be understood by noting that the production rate of RR increases with the density of ionized gas. At number densities  $n_{\text{H}} < 10^{-2} \text{ cm}^{-3}$ , where the gas is highly ionized, the photoionization rate due to recombination photons is proportional to the density (i.e.,  $\Gamma_{\text{RR}} \propto n_{\text{H}}$ ). At higher densities on the other hand, the gas becomes neutral. As a result, the density of ionized gas decreases with increasing density and the production rate of recombination photons decreases. Therefore, there is a peak in the photoionization rate due to RR around the self-shielding density. At very low densities, the superposition of recombination photons which have escaped from higher densities becomes dominant and the net photoionization rate of recombination photons flattens. Note that our simulations may underestimate this asymptotic rate because our simulation volumes are small compared to the mean free path for ionizing radiation (which is  $\sim 100$  Mpc at  $z \sim 3$ ). On the other hand, the neglect of cosmological redshifting for RR will result in overestimation of its photoionization rate on large scales. Recombination photons also leak from lower densities to self-shielded regions, smoothing the transition between highly ionized and highly neutral gas. At high densities, in the absence of the UVB ionizing photons, RR and collisional ionization can boost each other by providing more free electrons and ions.

In Figure 2.7 we compare hydrogen neutral fraction and photoionization rate profiles for different assumptions about RR. The hydrogen neutral fraction profile based on a precise RT calculation of RR is close to the Case A result at low densities ( $n_{\text{H}} \lesssim 10^{-3} \text{ cm}^{-3}$ ) but converges to the Case B result at high densities ( $n_{\text{H}} \gtrsim 10^{-1} \text{ cm}^{-3}$ ). This suggests that the neutral fraction profile, though not the ionization rate, can be modeled by switching from Case A to Case B recombination at  $n_{\text{H}} \sim n_{\text{H,SSH}}$  (e.g., Altay et al., 2011; McQuinn et al., 2011).

### 2.3.5 Evolution

The general trends in the profile of the photoionization rates with density and their influences on the distribution of HI are not very sensitive to redshift. However, as shown in table 2.2, the intensity and hardness of the UVB radiation change with redshift which, in turn, changes the self-shielding density. Moreover, as the Universe expands, the average density of absorbers decreases and their distributions evolve. The larger structures that form at lower redshifts drastically change the temperature structure of the gas at low and intermediate densities where collisional ionization becomes the dominant process. In the top-left panel of Figure 2.8, the evolution of the hydrogen neutral fraction is il-



**Figure 2.8:** Evolution of the hydrogen neutral fraction profile (*top-left*) and various ionization rates as a function of density. *Top-right*, *bottom-left* and *bottom-right* panels show, respectively, the total photoionization rates, collisional ionization rates and RR photoionization rates. All the H I fractions and collisional ionization rates which are sensitive to both collisional ionization and photoionization are taken from the *L50N512-W3* simulation. Photoionization rates at  $z \geq 2$  are based on the *L06N128* simulation. At lower redshifts (i.e.,  $z = 0$  and  $1$ ), where the box size become important because of the collisional ionization and its effect on changing the self-shielding, we used a representative sub-volume of the *L50N512-W3* simulation to calculate the photoionization rate profile. While the overall shape of the UVB photoionization rate profile is similar at different redshifts, the collisional ionization becomes increasingly stronger at lower redshifts and strongly reduces the hydrogen neutral fractions at densities  $n_{\text{H}} \lesssim 10^{-3} \text{ cm}^{-3}$ .

illustrated for the *L50N512-W3* simulation. As discussed in §2.3.3 and Appendix A1, since the photoionization rate profiles are converged with box size and resolution, we apply the profiles derived from a RT simulation of a smaller box, or a subset of the big box at lower redshifts<sup>8</sup>, to calculate the neutral fractions in this big box. Figure 2.8 shows that the neutral fraction profiles are similar in shape at high redshifts but that at  $z \leq 1$  the profiles are largely different, particularly at low hydrogen number densities, due to the evolving collisional ionization rates.

The evolution of the collisional ionization rate profiles is shown in the bottom-left panel of Figure 2.8. At  $z \geq 2$  and for  $n_{\text{H}} < 10^{-2} \text{ cm}^{-3}$ , the collisional ionization rate is not high enough to compete with the UVB photoionization rate. At lower redshifts and for number densities  $n_{\text{H}} \lesssim 3 \times 10^{-3} \text{ cm}^{-3}$ , on the other hand, collisional ionization dominates<sup>9</sup>. Indeed, the median collisional ionization rates are more than 100 times higher than the UVB photoionization rate at densities around the expected self-shielding thresholds. Collisional ionization therefore helps the UVB ionizing photons to penetrate to higher densities without being significantly absorbed. As a result, self-shielding starts at densities higher than expected from equation (2.13). The signature of collisional ionization on the hydrogen neutral fraction is more dramatic at low densities and partly compensates for the lower UVB intensity at  $z = 0$ . This results in a flattening of  $f(N_{\text{HI}}, z)$  at column densities  $N_{\text{HI}} \lesssim 10^{16} \text{ cm}^{-2}$  as shown in Figure 2.1.

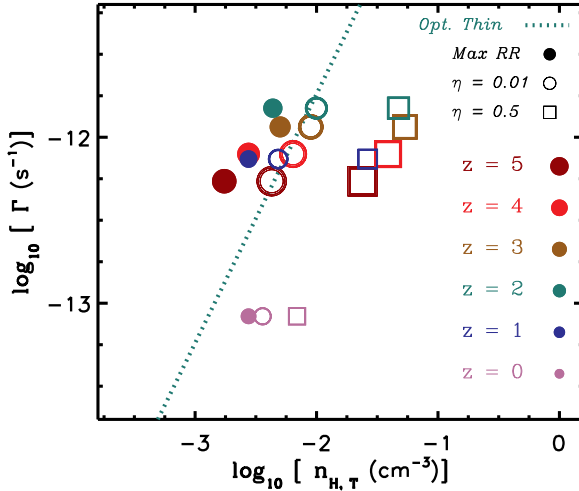
As mentioned above, at low redshifts (e.g.,  $z = 0$ ) the collisional ionization rate peaks at densities higher than the expected self-shielding threshold against the UVB. As a result, the total photoionization rate falls off rapidly together with the drop in the collisional ionization rate. Therefore, the drop in the hydrogen ionized fraction, and hence the resulting free electron density, is much sharper at lower redshifts. This causes a steeper high-density fall-off in the collisional ionization rate as shown in the bottom-left panel of Figure 2.8.

The differences between the total photoionization rates at different redshifts shown in the top-right panel of Figure 2.8, are caused by the evolution of the UVB intensity and its hardness, which affects the self-shielding density thresholds (see equation 2.13). On the other hand, as we showed in the previous section, the peak of the photoionization rate produced by RR tracks the self-shielding density. As a result the peak of the RR photoionization rate also changes with redshift as illustrated in the bottom-right panel of Figure 2.8).

The filled circles in Figure 2.9 indicate the UVB photoionization rate versus the number density at which the RR photoionization rate peaks. The self-shielding density expected from the Jeans scaling argument (equation 2.13) is also shown (green dotted line). The peaks in the RR photoionization rate in RT

<sup>8</sup>Strong collisional ionization at low redshifts can change the self-shielding. Therefore, for our RT simulation at low redshifts (i.e.,  $z \lesssim 1$ ) we used a representative sub-volume of the *L50N512-W3* simulation sufficiently large for the collisional ionization rates to be converged.

<sup>9</sup>We note that at redshift  $z \lesssim 1$  box sizes  $L_{\text{Box}} \gtrsim 25$  comoving  $h^{-1} \text{ Mpc}$  are required to fully capture the large-scale accretion shocks and to produce converged collisional ionization rates.



**Figure 2.9:** The photoionization from recombination radiation peaks at the expected self-shielding density and smooths the transition between highly ionized gas and self-shielded gas. This is illustrated by showing different characteristic densities for various redshifts. Filled circles show the density at which recombination radiation peaks, while open circles and squares show the H I number density corresponding to hydrogen neutral fractions of  $10^{-2}$  and 0.5, respectively. The self-shielding density threshold for a given photoionization rate expected from the Jeans scaling argument (equation 2.13) is also indicated by the green dotted line. The Jeans scaling argument works well, except at  $z = 0$  when collisional ionization is important.

simulations follow this expected scaling for  $z \geq 1$ . However, the  $z = 0$  result deviates from this trend since collisional ionization affects the self-shielding density threshold, a factor that is not captured by equation (2.13).

As a result of the RR photoionization rate peaking around the self-shielding density threshold, the transition between highly ionized and nearly neutral gas becomes more extended at all redshifts. To illustrate the smoothness of this transition, the densities at which the median hydrogen neutral fractions are  $10^{-2}$  and 0.5 are shown in Figure 2.9 with open circles and squares, respectively. The densities at which the hydrogen neutral fraction is  $10^{-2}$  are slightly higher than the densities at which the RR photoionization rate peaks (filled circles). The evolution agrees with the trend expected from the Jeans scaling argument and the self-shielding density. The exception is again  $z = 0$ , where the large collisional ionization rate at number densities  $n_{\text{H}} \sim 10^{-3} - 10^{-2} \text{ cm}^{-3}$  shifts the transition to neutral fraction of 0.5 to densities that are  $\approx 1$  dex higher. However, the relation between the photoionization rate and the density still follows the slope expected from equation (2.13).

It is interesting to note that the UVB spectral shape at  $z = 4$  is slightly harder than at  $z = 1$  while the UVB intensities at these two redshifts are similar. This

results in a deeper penetration of ionizing photons at  $z = 4$ . Consequently, the densities corresponding to the indicated neutral fractions (i.e.,  $10^{-2}$  and 0.5) at  $z = 1$  are lower than their counterparts at  $z = 4$ .

## 2.4 Conclusions

We combined a set of cosmological hydrodynamical simulations with an accurate RT simulation of the UVB radiation to compute the HI column density distribution function and its evolution. We ignored the effect of local sources of ionizing radiation, but we did include a self-consistent treatment of recombination radiation.

Our RT results for the distribution of photoionization rates at different densities are converged with respect to the simulation box size and resolution. Therefore, the resulting photoionization rate can be expressed as a function of the hydrogen density and the UVB. We provided a fit for the median total photoionization rate as a function of density that can be used with any desired UVB model to take into account the effect of HI self-shielding in cosmological simulations without the need to perform RT.

The CDDF,  $f(N_{\text{HI}}, z)$ , predicted by our RT simulations is in excellent agreement with observational constraints at all redshifts ( $z = 0 - 5$ ) and reproduces the slopes of the observed  $f(N_{\text{HI}}, z)$  function for a wide range of HI column densities. At low HI column densities, the CDDF is a steep function which decreases with increasing  $N_{\text{HI}}$  before it flattens at  $N_{\text{HI}} \gtrsim 10^{18} \text{ cm}^{-2}$  due to self-shielding. At  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-3}$  on the other hand,  $f(N_{\text{HI}}, z)$  is determined mainly by the intrinsic distribution of total hydrogen and the  $\text{H}_2$  fraction.

We showed that the  $N_{\text{HI}} - n_{\text{H}}$  relationship can be explained by a simple Jeans scaling. This argument assumes HI absorbers to be self-gravitating systems close to local hydrostatic equilibrium (Schaye, 2001) and to be either neutral or in photoionization equilibrium in the presence of an ionizing radiation field. However, at  $z = 0$  the analytic treatment underestimates the self-shielding density threshold due to its neglect of collisional ionization.

The high HI column density end of the predicted  $f(N_{\text{HI}}, z)$  evolves only weakly from  $z = 5$  to  $z = 0$ , consistent with observations. In the Lyman limit range of the distribution function, the slope of  $f(N_{\text{HI}}, z)$  remains the same at all redshifts. However, at  $z > 3$  the number of absorbers increases with redshift as the Universe becomes denser while the UVB intensity remains similar. At lower redshifts, on the other hand, the combination of a decreasing UVB intensity and the expansion of the Universe results in a non-evolving  $f(N_{\text{HI}}, z)$ . In contrast, the number of absorbers with lower HI column densities (i.e., the Ly $\alpha$  forest) decreases significantly from  $z \sim 3$ . We showed that this results in part from the stronger collisional ionization at redshifts  $z \lesssim 1$ , which compensates for the lower intensity of the UVB. The increasing importance of collisional ionization is due to the rise in the fraction of hot gas due to shock-heating associated with

the formation of structure.

The inclusion of diffuse recombination radiation smooths the transition between optically thin and thick gas. Consequently, the transition to highly neutral gas is not as sharp as what has been assumed in some previous works (e.g., Nagamine et al., 2010; Yajima et al., 2011; Goerdt et al., 2012). For instance, the difference in the gas density at which hydrogen is highly ionized (i.e.,  $n_{\text{HI}}/n_{\text{H}} \lesssim 0.01$ ) and the density at which gas is highly neutral (i.e.,  $n_{\text{HI}}/n_{\text{H}} \gtrsim 0.5$ ) is more than one order of magnitude (see Figure 2.9). As a result, assuming a sharp self-shielding density threshold at the density for which the optical depth of ionizing photons is  $\sim 1$ , overestimates the resulting neutral hydrogen mass by a factor of a few.

Our simulations adopted some commonly used approximations (e.g., neglecting helium RT effects, using a gray approximation in order to mimic the UVB spectra, neglecting absorption by dust and local sources of ionizing radiation). Our tests show that most of those approximations have negligible effects on our results. But there are some assumptions which require further investigation. For instance, the presence of young stars in high-density regions could change the H I CDDF, especially at high H I column densities through feedback and emission of ionizing photons. Indeed, we will show in Rahmati et al. (2013) that for very high column densities the ionizing radiation from young stars can reduce the  $f(N_{\text{HI}}, z)$  by 0.5-1 dex.

## Acknowledgments

We thank the anonymous referee for a helpful report. We thank Garbriel Altay for providing us with his simulation results and a compilation of the observed H I CDDF. We also would like to thank Kristian Finlator, J. Xavier Prochaska, Tom Theuns and all the members of the OWLS team for valuable discussions and Marcel Haas, Joakim Rosdahl, Maryam Shirazi and Freeke van de Voort for helpful comments on an earlier version of the paper this chapter is based on. The simulations presented here were run on the Cosmology Machine at the Institute for Computational Cosmology in Durham (which is part of the DiRAC Facility jointly funded by STFC, the Large Facilities Capital Fund of BIS, and Durham University) as part of the Virgo Consortium research programme and on Stella, the LOFAR BlueGene/L system in Groningen. This work was sponsored by the National Computing Facilities Foundation (NCF) for the use of supercomputer facilities, with financial support from the Netherlands Organization for Scientific Research (NWO), also through a VIDI grant and an NWO open competition grant. We also benefited from funding from NOVA, from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement 278594-GasAroundGalaxies and from the Marie Curie Training Network CosmoComp (PITN-GA-2009-238356). AHP receives funding from the European Union's Seventh Framework Programme

---

(FP7/2007-2013) under grant agreement number 301096.

## References

- Aguirre, A., Dow-Hygelund, C., Schaye, J., & Theuns, T. 2008, *ApJ*, 689, 851
- Altay, G., Theuns, T., Schaye, J., Crighton, N. H. M., & Dalla Vecchia, C. 2011, *ApJL*, 737, L37
- Blitz, L., & Rosolowsky, E. 2006, *ApJ*, 650, 933
- Cen, R., Ostriker, J. P., Prochaska, J. X., & Wolfe, A. M. 2003, *ApJ*, 598, 741
- Cen, R. 2012, *ApJ*, 748, 121
- Chabrier, G. 2003, *PASP*, 115, 763
- Dalla Vecchia, C., & Schaye, J. 2008, *MNRAS*, 387, 1431
- Davé, R., Oppenheimer, B. D., Katz, N., Kollmeier, J. A., & Weinberg, D. H. 2010, *MNRAS*, 408, 2051
- Duffy, A. R., Kay, S. T., Battye, R. A., et al. 2012, *MNRAS*, 420, 2799
- Erkal, D., Gnedin, N. Y., & Kravtsov, A. V. 2012, arXiv:1201.3653
- Faucher-Giguère, C.-A., Lidz, A., Zaldarriaga, M., & Hernquist, L. 2009, *ApJ*, 703, 1416
- Faucher-Giguère, C.-A., Kereš, D., Dijkstra, M., Hernquist, L., & Zaldarriaga, M. 2010, *ApJ*, 725, 633
- Friedrich, M. M., Mellema, G., Iliev, I. T., & Shapiro, P. R. 2012, *MNRAS*, 2385
- Fumagalli, M., Prochaska, J. X., Kasen, D., et al. 2011, *MNRAS*, 418, 1796
- Furlanetto, S. R., Schaye, J., Springel, V., & Hernquist, L. 2005, *ApJ*, 622, 7
- 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
- Gnedin, N. Y., Kravtsov, A. V., & Chen, H.-W. 2008, *ApJ*, 672, 765
- Goerdt, T., Dekel, A., Sternberg, A., Gnat, O., & Ceverino, D. 2012, arXiv:1205.2021
- Haardt F., Madau P., 2001, in *Clusters of Galaxies and the High Redshift Universe Observed in X-rays*, Neumann D. M., Tran J. T. V., eds.
- Haardt, F., & Madau, P. 2012, *ApJ*, 746, 125
- Haehnelt, M. G., Steinmetz, M., & Rauch, M. 1998, *ApJ*, 495, 647
- Hui, L., & Gnedin, N. Y. 1997, *MNRAS*, 292, 27
- Janknecht, E., Reimers, D., Lopez, S., & Tytler, D. 2006, *A&A*, 458, 427
- Katz, N., Weinberg, D. H., Hernquist, L., & Miralda-Escude, J. 1996, *ApJL*, 457, L57
- Kim, T.-S., Carswell, R. F., Cristiani, S., D’Odorico, S., & Giallongo, E. 2002, *MNRAS*, 335, 555
- Komatsu, E., et al. 2011, *ApJS*, 192, 18
- Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009a, *ApJ*, 693, 216
- Krumholz, M. R., Ellison, S. L., Prochaska, J. X., & Tumlinson, J. 2009b, *ApJL*, 701, L12
- Lehner, N., Savage, B. D., Richter, P., et al. 2007, *ApJ*, 658, 680



## REFERENCES

---

- McQuinn, M., & Switzer, E. R. 2010, *MNRAS*, 408, 1945
- McQuinn, M., Oh, S. P., & Faucher-Giguère, C.-A. 2011, *ApJ*, 743, 82
- Miralda-Escudé, J. 2005, *ApJL*, 620, L91
- Nagamine, K., Springel, V., & Hernquist, L. 2004, *MNRAS*, 348, 421
- Nagamine, K., Wolfe, A. M., Hernquist, L., & Springel, V. 2007, *ApJ*, 660, 945
- Nagamine, K., Choi, J.-H., & Yajima, H. 2010, *ApJL*, 725, L219
- Noterdaeme, P., Petitjean, P., Ledoux, C., & Srianand, R. 2009, *A&A*, 505, 1087
- Noterdaeme, P., Petitjean, P., Carithers, W. C., et al. 2012, arXiv:1210.1213
- O’Meara, J. M., Prochaska, J. X., Burles, S., et al. 2007, *ApJ*, 656, 666
- O’Meara, J. M., Prochaska, J. X., Worseck, G., Chen, H.-W., & Madau, P. 2012, arXiv:1204.3093
- Osterbrock, D. E., & Ferland, G. J. 2006, *Astrophysics of gaseous nebulae and active galactic nuclei*, 2nd. ed. by D.E. Osterbrock and G.J. Ferland. Sausalito, CA: University Science Books, 2006,
- Pawlik, A. H., & Schaye, J. 2008, *MNRAS*, 389, 651
- Pawlik, A. H., & Schaye, J. 2011, *MNRAS*, 412, 1943
- Péroux, C., Dessauges-Zavadsky, M., D’Odorico, S., Sun Kim, T., & McMahon, R. G. 2005, *MNRAS*, 363, 479
- Petitjean, P., Bergeron, J., & Puget, J. L. 1992, *A&A*, 265, 375
- Pontzen, A., Governato, F., Pettini, M., et al. 2008, *MNRAS*, 390, 1349
- Prochaska, J. X., & Wolfe, A. M. 2009, *ApJ*, 696, 1543
- Prochaska, J. X., Worseck, G., & O’Meara, J. M. 2009, *ApJL*, 705, L113
- Prochaska, J. X., O’Meara, J. M., & Worseck, G. 2010, *ApJ*, 718, 392
- Rahmati, A., & van der Werf, P. P. 2011, *MNRAS*, 418, 176
- Rahmati, A., Schaye, J., Pawlik, A. H., & Raičević, M. 2013, *MNRAS*, 431, 2261
- Razoumov, A. O., Norman, M. L., Prochaska, J. X., & Wolfe, A. M. 2006, *ApJ*, 645, 55
- Ribaudo, J., Lehner, N., & Howk, J. C. 2011, *ApJ*, 736, 42
- Robbins, D. 1978, *Amer. Math. Monthly* 85, 278
- Schaye, J. 2001a, *ApJ*, 559, 507
- Schaye, J. 2001b, *ApJL*, 562, 95
- Schaye, J. 2004, *ApJ*, 609, 667
- Schaye, J. 2006, *ApJ*, 643, 59
- Schaye, J., & Dalla Vecchia, C. 2008, *MNRAS*, 383, 1210
- Schaye, J., Dalla Vecchia, C., Booth, C. M., et al. 2010, *MNRAS*, 402, 1536
- Springel, V. 2005, *MNRAS*, 364, 1105
- Tepper-García, T., Richter, P., Schaye, J., Booth, C. M.; Dalla Vecchia, C.; Theuns, T. 2012, *MNRAS*, 425, 1640
- Tytler, D. 1987, *ApJ*, 321, 49
- Tajiri, Y., & Umemura, M. 1998, *ApJ*, 502, 59
- Theuns, T., Leonard, A., Efstathiou, G., Pearce, F. R., & Thomas, P. A. 1998, *MNRAS*, 301, 478
- van de Voort, F., Schaye, J., Altay, G., & Theuns, T. 2012, *MNRAS*, 421, 2809
- Verner, D. A., Ferland, G. J., Korista, K. T., & Yakovlev, D. G. 1996, *ApJ*, 465, 487



- Weingartner, J. C., & Draine, B. T. 2001, ApJ, 548, 296  
 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  
 Yajima, H., Choi, J.-H., & Nagamine, K. 2011, arXiv:1112.5691  
 Zheng, Z., & Miralda-Escudé, J. 2002, ApJL, 568, L71  
 Zwaan, M. A., van der Hulst, J. M., Briggs, F. H., Verheijen, M. A. W., & Ryan-Weber, E. V. 2005, MNRAS, 364, 1467

## Appendix A: Photoionization rate as a function of density

### A1: Replacing the RT simulations with a fitting function

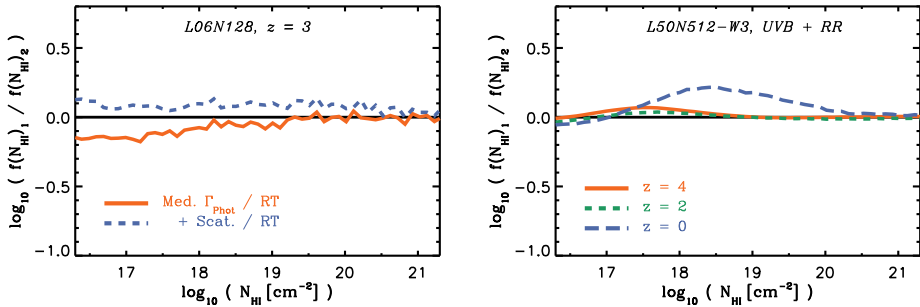
In §2.3.3 we demonstrated that the median of the simulated relation between the total photoionization rate,  $\Gamma_{\text{Phot}}$ , and density is converged with respect to resolution and box size. We used this result and provided fits to the median of this relation. We have exploited these fits to compute the neutral hydrogen fraction in cosmological simulations under the assumption of ionization equilibrium (see Appendix A2), without performing the computationally demanding RT. In this section, we discuss the accuracy of these fits.

The left panel of Figure 2.10 shows that using the median photoionization rates produces an H $\alpha$  CDDF in very good agreement with the H $\alpha$  CDDF obtained from the corresponding RT simulation (orange solid curve) at  $N_{\text{HI}} \gtrsim 10^{18} \text{ cm}^{-2}$ . However, there is a small systematic difference at lower column densities. One may think that this small difference is caused by the loss of information contained in the scatter in the photoionization rates at fixed density. We tested this hypothesis by including a log-normal random scatter around the median photoionization rate consistent with the scatter exhibited by the RT result. However, after accounting for the random scatter, the  $f(N_{\text{HI}}, z)$  is slightly overproduced compared to the full RT result at nearly all H $\alpha$  column densities.

We exploit the insensitivity of the shape of the  $\Gamma_{\text{Phot}}$ -density relation to the redshift, and propose the following fit to the photoionization rate,  $\Gamma_{\text{Phot}}$ ,

$$\frac{\Gamma_{\text{Phot}}}{\Gamma_{\text{UVB}}} = (1 - f) \left[ 1 + \left( \frac{n_{\text{H}}}{n_0} \right)^\beta \right]^{\alpha_1} + f \left[ 1 + \frac{n_{\text{H}}}{n_0} \right]^{\alpha_2}, \quad (2.15)$$

where  $\Gamma_{\text{UVB}}$  is the photoionization rate due to the ionizing background, and  $n_0$ ,  $\alpha_1$ ,  $\alpha_2$ , and  $\beta$  are parameters of the fit. The best-fit values of these parameters are listed in Table 2.3 and the photoionization rate-density relations they produce are compared with the RT simulations at redshifts  $z = 0$  and  $z = 4$  in Figure 2.11. At all redshifts, the best-fit value of  $n_0$  is almost identical to the self-shielding density threshold,  $n_{\text{H,SSH}}$ , defined in equation (2.13), and the characteristic slopes



**Figure 2.10:** *Left:* The ratio between the H I CDDF calculated using the RT based  $\Gamma_{\text{Phot}} - n_{\text{H}}$  relationship and the actual RT results for the *L06N128* simulation in the presence of the UVB and diffuse recombination radiation, at  $z = 3$ . The orange solid line shows the result of using the median photoionization rate-density profile predicted by the RT simulation. The blue dashed curve shows the result of including the scatter around the median in the calculations. *Right:* H I CDDFs calculated using the  $\Gamma_{\text{Phot}} - n_{\text{H}}$  fitting function (i.e., equation 2.14) are compared to the H I CDDFs for which the actual  $\Gamma_{\text{Phot}}$ -density relation from the RT simulations are used. Blue and green curves are for  $z = 0$  and  $z = 2$  respectively and the red curve is for  $z = 4$ . The difference between the RT result and the result of using the fitting function at  $z = 0$  is due to the importance of collisional ionization at  $z = 0$ . To capture this effect and to reproduce the RT results at  $z = 0$ , we advise using the best-fit parameters shown in Table 2.4. All the CDDFs are for the *L50N512-W3* simulation and in the presence of the UVB and diffuse recombination radiation.

of the photoionization rate-density relation are similar. This suggest that one can find a single set of best-fit values to reproduce the RT results at  $z \gtrsim 1$ . The corresponding best-fit parameter values are (see also equation 2.14)  $\alpha_1 = -2.28 \pm 0.31$ ,  $\alpha_2 = -0.84 \pm 0.11$ ,  $n_0 = (1.003 \pm 0.005) \times n_{\text{H,SSH}}$ ,  $\beta = 1.64 \pm 0.19$  and  $f = 0.02 \pm 0.0089$ .

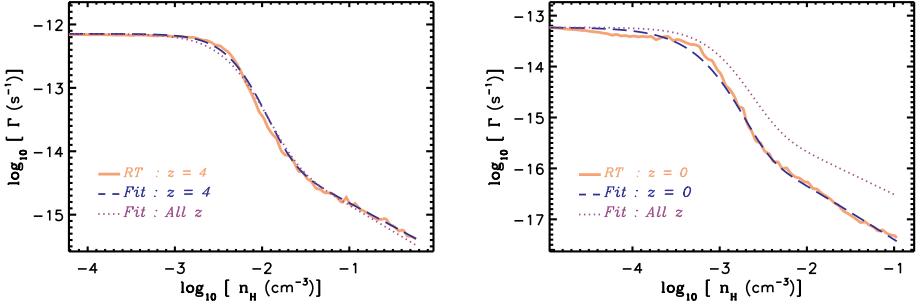
In the right panel of Figure 2.10, the ratio between the H I CDDF calculated using the fitting function presented in equation (2.14) and the RT based H I CDDF (i.e., calculated using the median of the photoionization rate-density relation in the RT simulations) is shown for the *L50N512-W3* and at  $z = 0, 2$  and 4. This illustrates that the fitting function reproduces the RT results accurately, except at  $z = 0$ . As explained in §2.3, this is expected since at low redshifts collisional ionization affects the self-shielding and the resulting photoionization rate-density profile. However, a separate fit can be obtained using converged RT results at  $z = 0$ . The parameters that define such a fit are shown in Table 2.4.

## A2: The equilibrium hydrogen neutral fraction

In this section we explain how to derive the neutral fraction in ionization equilibrium. Equating the total number of ionizations per unit time per unit volume

**Table 2.3:** The best-fit parameters for equation (2.15) at different redshifts based on RT results in the *L06N128* simulation.

Redshift	$\log [n_0] (\text{cm}^{-3})$	$\alpha_1$	$\alpha_2$	$\beta$	$1 - f$
$z = 1-5$	$\log [n_{\text{H,SSH}}]$	-2.28	-0.84	1.64	0.98
$z = 0$	-2.94	-3.98	-1.09	1.29	0.99
$z = 1$	-2.29	-2.94	-0.90	1.21	0.97
$z = 2$	-2.06	-2.22	-1.09	1.75	0.97
$z = 3$	-2.13	-1.99	-0.88	1.72	0.96
$z = 4$	-2.23	-2.05	-0.75	1.93	0.98
$z = 5$	-2.35	-2.63	-0.57	1.77	0.99



**Figure 2.11:** Comparisons between the total photoionization rates as a function of density in the *L06N128* simulation. Photoionization rates based on the RT simulations and best-fit functions at  $z = 4$  and  $z = 0$  are shown in the *left* and *right* panels, respectively. In each panel, the RT result is shown with the orange solid curve. The best fit to the RT result at a given redshift (equation 2.15 and Table 2.3) is shown with the blue dashed curve and the best fit to the RT results at  $z = 1 - 5$  (equation 2.14) is shown with the purple dotted curve. As shown in the *right* panel, because of the impact of collisional ionization on self-shielding, the low redshift photoionization curve (the blue dashed curve) deviates from the best fit to the results at higher redshifts (the purple dotted curve). To resolve this issue and to capture the impact of collisional ionization, we advise using the best-fit parameters shown in Table 2.4 for  $z = 0$ .

**Table 2.4:** The best-fit parameters for equation (2.15) at  $z = 0$  based on RT results for the *L50N512* simulation. To capture the impact of collisional ionization on the self-shielding, one needs to use large cosmological simulations. The simulation with a box size of  $50 h^{-1}\text{Mpc}$  results in converged collisional ionizations.

Redshift	$\log [n_0] (\text{cm}^{-3})$	$\alpha_1$	$\alpha_2$	$\beta$	$1 - f$
$z = 0$	-2.56	-1.86	-0.51	2.83	0.99

with the total number of recombinations per unit time per unit volume, we obtain

$$n_{\text{HI}} \Gamma_{\text{TOT}} = \alpha_A n_e n_{\text{HII}}, \quad (2.16)$$

where  $n_{\text{HI}}$ ,  $n_e$  and  $n_{\text{HII}}$  are the number densities of neutral hydrogen atoms, free electrons and protons, respectively.  $\Gamma_{\text{TOT}}$  is the total ionization rate per neutral hydrogen atom and  $\alpha_A$  is the Case A recombination rate<sup>10</sup> for which we use the fitting function given by Hui & Gnedin (1997):

$$\alpha_A = 1.269 \times 10^{-13} \frac{\lambda^{1.503}}{\left(1 + (\lambda/0.522)^{0.47}\right)^{1.923}} \text{ cm}^3 \text{ s}^{-1}, \quad (2.17)$$

where  $\lambda = 315614/T$ .

Defining the hydrogen neutral fraction as the ratio between the number densities of neutral hydrogen and total hydrogen,  $\eta = n_{\text{HI}}/n_{\text{H}}$ , and ignoring helium (which is an excellent approximation, see Appendix D2), we can rewrite equation (2.17) as:

$$\eta \Gamma_{\text{TOT}} = \alpha_A (1 - \eta)^2 n_{\text{H}}. \quad (2.18)$$

Furthermore, we can assume that the total ionization rate,  $\Gamma_{\text{TOT}}$ , consists of two components: the total photoionization rate,  $\Gamma_{\text{Phot}}$ , and the collisional ionization rate,  $\Gamma_{\text{Col}}$ :

$$\Gamma_{\text{TOT}} = \Gamma_{\text{Phot}} + \Gamma_{\text{Col}}, \quad (2.19)$$

where  $\Gamma_{\text{Col}} = \Lambda_T (1 - \eta) n_{\text{H}}$ . The photoionization rate can be expressed as a function of density using equation (2.14). For  $\Lambda_T$ , which depends only on temperature, we use a relation given in Theuns et al. (1998):

$$\Lambda_T = 1.17 \times 10^{-10} \frac{T^{1/2} \exp(-157809/T)}{1 + \sqrt{T/10^5}} \text{ cm}^3 \text{ s}^{-1}. \quad (2.20)$$

We can now rearrange equation (2.18) as a quadratic equation:

$$A \eta^2 - B \eta + C = 0, \quad (2.21)$$

with  $A = \alpha_A + \Lambda_T$ ,  $B = 2\alpha_A + \frac{\Gamma_{\text{Phot}}}{n_{\text{H}}} + \Lambda_T$  and  $C = \alpha_A$  which gives:

$$\eta = \frac{B - \sqrt{B^2 - 4AC}}{2A}. \quad (2.22)$$

Using the last equation one can calculate the equilibrium hydrogen neutral fraction for a given  $n_{\text{H}}$  and temperature.

---

<sup>10</sup>The use of Case B is more appropriate for  $n_{\text{H}} > n_{\text{H,SSH}}$ . However, we assume the photoionization due to RR is included in  $\Gamma_{\text{TOT}}$ , e.g., by using the best-fit function that is presented in equation 2.14. Therefore, Case A recombination should be adopted even at high densities.

## Appendix B: The effects of box size, cosmological parameters and resolution on the HI CDDF

The size of the simulation box may limit the abundance and the density of the densest systems captured by the simulation. In other words, very massive structures, which may be associated with the highest HI column densities, cannot be formed in a small cosmological box. Indeed, as shown in the top panels of Figure 2.12, one needs to use cosmological boxes larger than  $\gtrsim 25$  comoving  $h^{-1}\text{Mpc}$  in order to achieve convergence in the HI distribution (see also Altay et al., 2011). On the other hand, the bottom-right panel of Figure 2.12 shows that changing the resolution of the cosmological simulations also affects  $f(N_{\text{HI}}, z)$ , although the effect is small.

The adopted cosmological parameters also affect the gas distribution and hence the HI CDDF. For instance, one expects that the number of absorbers at a given density varies with the density parameter  $\Omega_b$ , and the root mean square amplitude of density fluctuations  $\sigma_8$ . The bottom-left panel of Figure 2.12 shows the ratio of column densities in simulations assuming WMAP 7-year and 3-year parameters. The ratio is only weakly dependent on the box size of the simulation and its resolution. This motivates us to use this ratio to convert the HI CDDF between the two cosmologies for all box sizes and resolutions (at any given redshift). While this is an approximate way of correcting for the difference in the cosmological parameters, it does not affect the main conclusions presented in this work (e.g., the lack of evolution of  $f(N_{\text{HI}}, z)$ ).

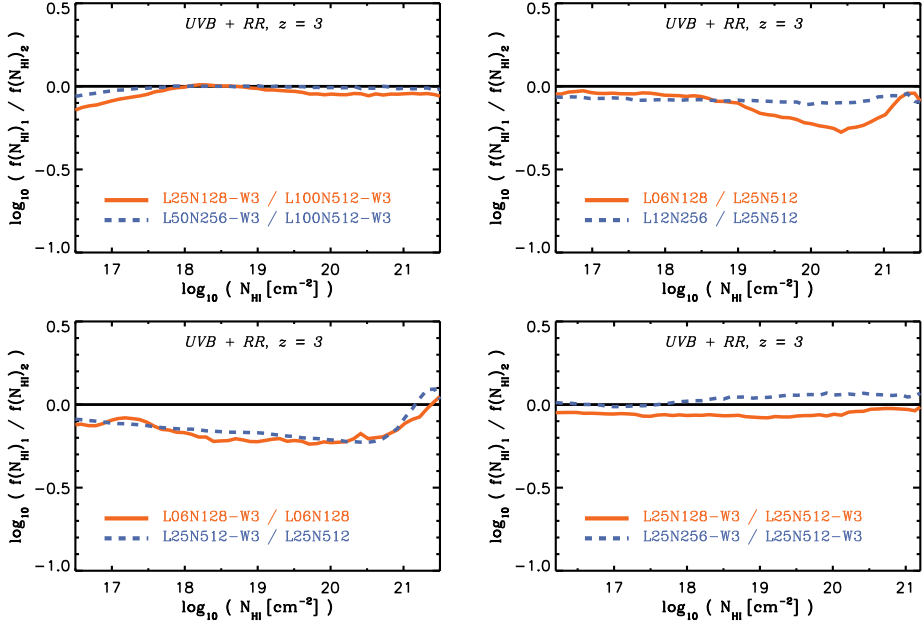
## Appendix C: RT convergence tests

### C1: Angular resolution

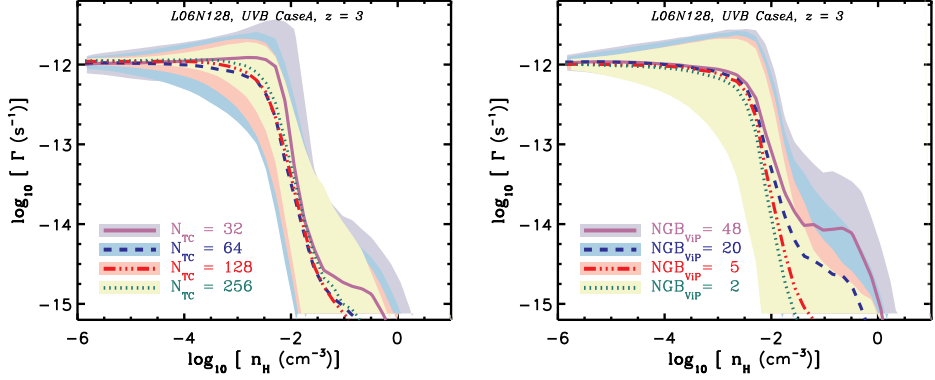
The left-hand panel of Figure 2.13 shows the dependence of photoionization rates on the adopted angular resolution, i.e., the opening angle of the transmission cones  $4\pi/N_{\text{TC}}$ . The photoionization rates are converged for  $N_{\text{TC}} = 64$  (our fiducial value) or higher.

### C2: The number of ViP neighbors

The right panel of Figure 2.13 shows the dependence of the photoionization rates on the number of SPH neighbors of ViPs. As discussed in §2.2.2, ViPs distribute the ionizing photons they absorb among their  $\text{NGB}_{\text{ViP}}$  nearest SPH neighbors. The larger the number of neighbors, the larger the volume over which photons are distributed, and the more extended is the transition between highly ionized and self-shielded gas. The photoionization rates converge for  $\lesssim 5$  ViP neighbors (our fiducial value is 5).



**Figure 2.12:** The relative changes in the H I CDDF using different resolutions, box sizes and cosmologies in the presence of the UVB and diffuse recombination radiation. The *top-left* panel shows the effect of box size on  $f(N_{\text{HI}}, z)$  for a fixed resolution at  $z = 3$ , where the orange solid (blue dashed) curve shows the difference between using a box size of  $L = 25$  (50) comoving  $h^{-1}\text{Mpc}$  and a box size of  $L = 100$  comoving  $h^{-1}\text{Mpc}$ . The *top-right* panel shows the same effect but for smaller box sizes: the orange solid (blue dashed) curve shows the difference between using a box size of  $L = 6$  (12) comoving  $h^{-1}\text{Mpc}$  and a box size of  $L = 25$  comoving  $h^{-1}\text{Mpc}$ . The *bottom-left* shows the effect of using a cosmology consistent with WMAP 3-year results instead of using a cosmology based on the WMAP 7-year constraints. The orange solid and blue dashed curves show this effect for simulations with box sizes of  $L = 6$  and 25 comoving  $h^{-1}\text{Mpc}$ , respectively. The *bottom-right* panel shows the effect of resolution.



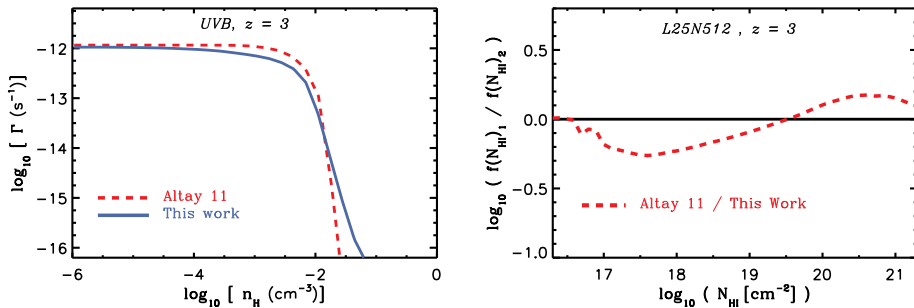
**Figure 2.13:** The UVB photoionization rate is converged for our adopted angular resolution, i.e.,  $N_{\text{TC}} = 64$ , as shown in the *left* panel and our adopted number of ViP neighbors, i.e.,  $N_{\text{GB\_ViP}} = 5$ , as shown in the *right* panel. Photoionization rate profiles are shown for the *L06N128* simulation in the presence of the UVB radiation where the Case A recombination is adopted. The curves show the medians and the shaded areas around them indicate the 15% – 85% percentiles.

### C3: Direct comparison with another RT method

Altay et al. (2011) used cosmological simulations from the reference model of the OWLS project (Schaye et al., 2010), i.e., a simulation run with the same hydro code as we used in this work, to investigate the effect of the UVB on the  $\text{H I}$  CDDF at  $z = 3$ . However, they employed a ray-tracing method very different from the RT method we use here. Furthermore, they did not explicitly treat the transfer of recombination radiation. In Figure 2.14, we compare one of our UVB photoionization rate profiles<sup>11</sup> with the photoionization rate found by Altay et al. (2011) in a similar simulation. The overall agreement is very good, but the comparison also reveals important differences.

Altay et al. (2011) calculate the average optical depth around every SPH particle within a distance of 100 proper kpc, assuming the UVB is unattenuated at larger distances. Then, they use this optical depth to calculate the attenuation of the UVB photoionization rate for every particle. This procedure may underestimate the small but non-negligible absorption of UVB ionizing photons on large scales. Indeed, by tracing the self-consistent propagation of photons inside the simulation box, we have found that the UVB photoionization rate decreases gradually with increasing density up to the density of self-shielding. However, we note that the small differences between our UVB photoionization rates and those calculated by Altay et al. (2011) at densities below the self-shielding, become slightly smaller by increasing the angular resolution in our RT calculations

<sup>11</sup>Note that in our simulations the UVB photoionization rate is converged with the box size and the resolution as shown in §2.3.3.



**Figure 2.14:** *Left:* Median UVB photoionization rate as a function of density at  $z = 3$  using different RT methods. The red dashed curve shows the results based on the method that has been used in Altay et al. (2011) and the blue solid curve shows the result of this work. *Right:* The H I CDDF of the L25N512 simulation at  $z = 3$  using different RT methods and without RR. The red dashed curve shows the ratio between the H I CDDF given in Altay et al. (2011) and our results. This comparison shows that despite the overall agreement between our results and Altay et al. (2011), there are some important differences.

(see the left panel of Figure 2.13).

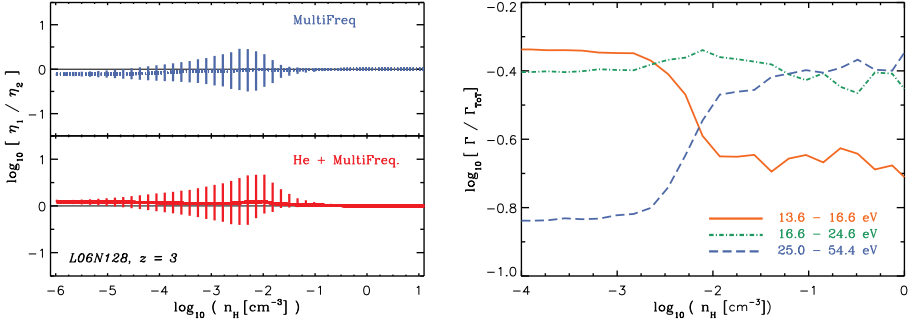
## Appendix D: Approximated processes

### D1: Multifrequency effects

As discussed in §2.2.3, in our RT simulations we have treated the multifrequency nature of the UVB radiation in the gray approximation (see equation 2.4). This approach does not capture the spectral hardening which is a consequence of variation of the absorption cross-sections with frequencies. We tested the impact of spectral hardening on the H I fractions by repeating the L06N128 simulation at  $z = 3$  with the UVB using 3 frequency bins. We used energy intervals  $[13.6 - 16.6]$ ,  $[16.6 - 24.6]$  and  $[24.6 - 54.4]$  eV and assumed that photons with higher frequencies are absorbed by He. The result is illustrated in the top section of the left panel in Figure 2.15, by plotting the ratio between the resulting hydrogen neutral fraction,  $\eta$ , and the same quantity in the original simulation that uses the gray approximation. This comparison shows that the simulation that uses multifrequency predicts hydrogen neutral fractions  $< 10\%$  lower at low densities (i.e.,  $n_{\text{H}} \lesssim 10^{-4} \text{ cm}^{-3}$ ). This does not change the resulting  $f(N_{\text{HI}}, z)$  noticeably at the column densities of interest here.

The spectral hardening captured in the simulation with 3 frequency bins is illustrated in the right panel of Figure 2.15. This figure shows the fractional contribution of different frequencies to the total UVB photoionization rate as a function of density. The red solid curve shows the contribution of the bin with





**Figure 2.15:** Spectral hardening and multifrequency treatment do not change the H I distribution significantly. *Left:* The ratio between hydrogen neutral fractions,  $\eta$ , obtained by using 3 frequency bins and by using the gray approximation is shown in the top-left panel. The ratio between hydrogen neutral fractions resulting from a simulation with 4 frequency bins and explicit He treatment and the same quantity using the gray approximation and without explicit He treatment is shown in the bottom-left panel. The vertical lines with different lengths indicate the 15% – 85% percentiles. *Right:* The fractional contribution of different frequency bins to the total UVB photoionization rate for the simulations with 3 frequency bins. All the RT calculations are performed using the *L06N128* simulation at  $z = 3$  in the presence of the UVB and assuming Case A recombination.

the lowest frequency and drops at the self-shielding density threshold. On the other hand, the fractional contribution of the hardest frequency bin increases at higher densities, as shown with the blue dashed curve. Despite the differences in the fractional contributions to the total UVB photoionization rate, the absolute photoionization rates drop rapidly at densities higher than the self-shielding threshold for all frequency bins.

## D2: Helium treatment

A simplifying assumption frequently used in RT simulations which aim to calculate the distribution of neutral hydrogen is to ignore helium in the ionization processes (e.g., Faucher-Giguère et al., 2009; McQuinn & Switzer, 2010; Altay et al., 2011). We adopted the same assumption in our RT calculations which implies that we implicitly assumed the ionization state of neutral helium and its interaction with free electrons to be similar to the trends followed by neutral hydrogen. This has been shown to be a good assumption (Osterbrock & Ferland, 2006; McQuinn & Switzer, 2010; Friedrich et al., 2012). Nevertheless, we tested the validity of our approximate helium treatment by repeating the *L06N128* simulation at  $z = 3$  with the UVB using 4 frequency bins and an explicit He treatment. The first three frequency bins are identical to the bins used in the previous section (i.e., [13.6 – 16.6], [16.6 – 24.6] and [24.6 – 54.4] eV) and the last bin is chosen to cover higher frequencies which are capable of He II ionization.

We adopted a helium mass fraction of 25% and a Case A recombination rate. The ratio between the resulting hydrogen neutral fraction and the same quantity when a single frequency is used and helium is not treated explicitly is illustrated in the bottom-left panel of Figure 2.15. The hydrogen neutral fractions are very close in the two simulations. However, the simulation with multifrequency and explicit He treatment results in hydrogen neutral fractions that are  $< 10\%$  higher at low densities (i.e.,  $n_{\text{H}} \lesssim 10^{-4} \text{ cm}^{-3}$ ). This difference is barely noticeable in the comparison between the two H<sub>I</sub> CDDFs (not shown).

# 3

## THE IMPACT OF LOCAL STELLAR RADIATION ON THE HI COLUMN DENSITY DISTRIBUTION

It is often assumed that local sources of ionizing radiation have little impact on the distribution of neutral hydrogen in the post-reionization Universe. While this is a good assumption for the intergalactic medium, analytic arguments suggest that local sources may typically be more important than the meta-galactic background radiation for high column density absorbers ( $N_{\text{HI}} > 10^{17} \text{ cm}^{-2}$ ). We post-process cosmological, hydrodynamical simulations with accurate radiation transport to investigate the impact of local stellar sources on the column density distribution function of neutral hydrogen. We demonstrate that the limited numerical resolution and the simplified treatment of the interstellar medium (ISM) that are typical of the current generation of cosmological simulations provide significant challenges, but that many of the problems can be overcome by taking two steps. First, using ISM particles rather than stellar particles as sources results in a much better sampling of the source distribution, effectively mimicking higher-resolution simulations. Second, by rescaling the source luminosities so that the amount of radiation escaping into the intergalactic medium agrees with that required to produce the observed background radiation, many of the results become insensitive to errors in the predicted fraction of the radiation that escapes the immediate vicinity of the sources. By adopting this strategy and by drastically varying the assumptions about the structure of the unresolved ISM, we conclude that we can robustly estimate the effect of local sources for column densities  $N_{\text{HI}} \ll 10^{21} \text{ cm}^{-2}$ . However, neither the escape fraction of ionizing radiation nor the effect of local sources on the abundance of  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$  systems can be predicted with confidence. We find that local stellar radiation is unimportant for  $N_{\text{HI}} \ll 10^{17} \text{ cm}^{-2}$ , but that it can affect Lyman Limit and Damped Ly $\alpha$  systems. For  $10^{18} < N_{\text{HI}} < 10^{21} \text{ cm}^{-2}$  the impact of local sources increases with redshift. At  $z = 5$  the abundance of absorbers is substantially reduced for  $N_{\text{HI}} \gg 10^{17} \text{ cm}^{-2}$ , but at  $z = 0$  the effect only becomes significant for  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ .

Alireza Rahmati, Joop Schaye, Andreas H. Pawlik, Milan Raičević  
*Monthly Notices of the Royal Astronomical Society*  
Volume 431, Issue 3, pp. 2261-2277 (2013)

### 3.1 Introduction

After the reionization of the Universe at  $z \gtrsim 6$ , hydrogen residing in the intergalactic medium (IGM) is kept highly ionized primarily by the meta-galactic UV background (UVB). The UVB is the integrated radiation that has been able to escape from sources into the IGM. Because the mean free path of ionizing photons is large compared to the scale at which the sources of ionizing radiation cluster, the UVB is expected to be close to uniform in the IGM. However, close to galaxies, the radiation field is dominated by local sources and hence more inhomogeneous.

Observations of neutral hydrogen (HI) in the Ly $\alpha$  forest mostly probe the low-density IGM which is typically far from star-forming regions. The statistical properties of the Ly $\alpha$  forest are therefore insensitive to the small-scale fluctuations in the UVB (e.g., Zuo, 1992; Croft, 2004). On the other hand, the neutral hydrogen in Damped Ly $\alpha$  (DLA; i.e.,  $N_{\text{HI}} > 10^{20.3} \text{ cm}^{-2}$ ) and Lyman Limit systems (LL; i.e.,  $10^{17.2} < N_{\text{HI}} \leq 10^{20.3} \text{ cm}^{-2}$ ), which are thought to originate inside or close to galaxies, might be substantially affected by radiation from local sources that are stronger than the ambient UVB (Gnedin, 2010). As a result, the abundances of the high HI column densities may also change significantly by locally produced radiation.

Indeed, Schaye (2006) and Miralda-Escudé (2005) have used analytic arguments to show that the impact of local radiation may be substantial for LL and (sub-)DLA systems, but should generally be very small at lower column densities. However, relatively little has been done to go beyond idealized analytic arguments and to simulate the effect of local radiation by taking into account the inhomogeneous distribution of sources and gas in and around galaxies. One of the main reasons for this is the computational expense of radiative transfer (RT) calculations in simulations with large numbers of sources. In addition, high resolution is required to capture the distribution of gas on small scales accurately. Because of these difficulties, most simulations of the cosmological HI distribution have ignored the impact of local radiation and focused only on the effect of the UVB (e.g., Katz et al., 1996; Gardner et al., 1997; Haehnelt et al., 1998; Cen et al., 2003; Nagamine et al., 2004; Razoumov et al., 2006; Pontzen et al., 2008; Altay et al., 2011; McQuinn et al., 2011; van de Voort et al., 2012; Rahmati et al., 2013; Bird et al., 2013). A few studies have taken into account local stellar radiation but their results are inconclusive: while Nagamine et al. (2010) and Yajima et al. (2012) found that local stellar radiation has a negligible impact on the distribution of HI, Fumagalli et al. (2011) found that the HI column density distribution above the Lyman limit is reduced by  $\sim 0.5$  dex due to local stellar radiation.

In this chapter, we investigate the impact of local stellar radiation on the HI distribution by combining cosmological hydrodynamical simulations with accurate RT. We find that the inclusion of local sources can dramatically change the predicted abundance of strong DLA systems and, depending on redshift,

also those of LL and weak DLA systems. On the other hand, lower H I column densities are hardly affected. We also show that resolution effects have a major impact. For instance, the resolution accessible to current cosmological simulations is insufficient to resolve the interstellar medium (ISM) on the scales relevant for the propagation of ionizing photons. The limited resolution also affects the source distribution which can change the resulting H I distribution, especially in low-mass galaxies. On top of that, assumptions about the structure of the unresolved multiphase ISM significantly affect the escape of stellar radiation into the IGM. Therefore, any attempt to use cosmological simulations to investigate the impact of local stellar radiation on the H I distribution may suffer from serious numerical artifacts.

Some of these difficulties can be circumvented by tuning the luminosities of the sources such that the escaped radiation can account for the observed UVB. Then the interaction between the radiation that reaches the IGM and the intervening gas can be studied on scales that are properly resolved in the simulations. We adopt this procedure to generate the observed UVB for various ISM models but find that our fiducial simulation reproduces the observed UVB without any tuning.

Among the known sources of radiation, quasars and massive stars are the most efficient producers of hydrogen ionizing photons and are therefore thought to be the main contributors to the UVB. Star-forming galaxies however are thought to be the dominant producers of the UVB at  $z \gtrsim 3$  (e.g., Haehnelt et al., 2001; Bolton et al., 2005; Faucher-Giguère et al., 2008a). Therefore, we only account for local radiation that is produced by star formation in our simulations. Moreover, we assume that the ionizing emissivity of baryons strictly follows the star formation rate. We use star-forming gas particles rather than young stellar particles as ionizing sources. Since the gas consumption time scale in the ISM is much larger than the lifetime of massive stars ( $\sim 10^9$  yr vs.  $\sim 10^7$  yr), there are many more star-forming gas particles than there are stellar particles young enough to efficiently emit ionizing radiation. Therefore, using star-forming gas particles as ionizing sources allows us to sample the source distribution better and hence to reduce the impact of the limited resolution of cosmological simulations.

The structure of the chapter is as follows. In §3.2 we use the observed relation between the star formation rate and gas surface densities to provide an analytic estimate for the photoionization rate that is produced by young stars in a uniform and optically thick ISM that is in good agreement with our simulation results. In §3.3 we discuss the details of our numerical simulations and RT calculations. The simulation results are presented in §3.4. We show how the local stellar radiation can generate the observed UVB. We also discuss the effect of unresolved ISM on the escape fraction of ionizing radiation and we investigate the impact of local stellar radiation on the H I column density distribution. Finally, we conclude in §3.5.

## 3.2 Photoionization rate in star-forming regions

A strong correlation between gas surface density and star formation rate has been observed in low-redshift galaxies (e.g., Kennicutt, 1998; Bigiel et al., 2008). In principle, such a relation can be combined with simplified assumptions to derive a typical photoionization rate that is expected from stellar radiation. In this section, we use this approach to estimate the average ionization rate that is produced by star formation as a function of gas (surface) density on galactic scales.

Assuming a Chabrier (2003) initial mass function (IMF), the observed Kennicutt-Schmidt law provides a relation between gas surface density,  $\Sigma_{\text{gas}}$ , and star formation rate surface density,  $\dot{\Sigma}_*$ , on kilo-parsec scales (Kennicutt, 1998):

$$\dot{\Sigma}_* \approx 1.5 \times 10^{-4} \text{ M}_\odot \text{ yr}^{-1} \text{ kpc}^{-2} \left( \frac{\Sigma_{\text{gas}}}{1 \text{ M}_\odot \text{ pc}^{-2}} \right)^{1.4}. \quad (3.1)$$

Equation (3.1) can be used to derive a relation between ionizing emissivity (per unit area),  $\dot{\Sigma}_\gamma$ , and the gas surface density. As we will discuss in §3.3.4, stellar population synthesis models indicate that the typical number of ionizing photons produced per unit time by a constant star formation rate is

$$\dot{Q}_\gamma \sim 2 \times 10^{53} \text{ s}^{-1} \left( \frac{\text{SFR}}{1 \text{ M}_\odot \text{ yr}^{-1}} \right). \quad (3.2)$$

Furthermore, as both models and observations suggest, the escape fraction of ionizing photons from galaxies is  $\ll 1$  (e.g., Shapley et al., 2006; Gnedin et al., 2008; Vanzella et al., 2010; Yajima et al., 2012; Paardekooper et al., 2011; Kim et al., 2012). This allows us to assume that most of the ionizing photons that are produced by star-forming gas are absorbed on scales  $\leq$  kpc. Therefore, the hydrogen photoionization rate, on galactic scales, can be computed:

$$\Gamma_* = \frac{\dot{\Sigma}_\gamma}{N_{\text{H}}} = \frac{\dot{Q}_\gamma \dot{\Sigma}_*}{N_{\text{H}}}, \quad (3.3)$$

where  $N_{\text{H}}$  is the hydrogen column density which can be obtained from the gas surface density:

$$N_{\text{H}} \approx 9.4 \times 10^{19} \text{ cm}^{-2} \left( \frac{\Sigma_{\text{gas}}}{1 \text{ M}_\odot \text{ pc}^{-2}} \right) \left( \frac{X}{0.75} \right), \quad (3.4)$$

where  $X$  is the hydrogen mass fraction. After assuming  $X = 0.75$  and substituting equations (3.1), (3.2) and (3.4) in equation (3.3), one gets

$$\Gamma_* \sim 8.5 \times 10^{-14} \text{ s}^{-1} \left( \frac{N_{\text{H}}}{10^{21} \text{ cm}^{-2}} \right)^{0.4}. \quad (3.5)$$

If the scale height of the disk is similar to the local Jeans scale, the hydrogen column density can be computed as a function of the hydrogen number density

(Schaye, 2001, 2004; Schaye & Dalla Vecchia, 2008)

$$N_{\text{H}} \sim 2.8 \times 10^{21} \text{ cm}^{-2} \left( \frac{n_{\text{H}}}{1 \text{ cm}^{-3}} \right)^{1/2} T_4^{1/2} \left( \frac{f_{\text{g}}}{f_{\text{th}}} \right)^{1/2}, \quad (3.6)$$

where  $T_4 \equiv T/10^4$  K,  $f_{\text{g}}$  is the total mass fraction in gas and  $f_{\text{th}}$  is the fraction of the pressure that is thermal (i.e.,  $P_{\text{th}} = f_{\text{th}} P_{\text{TOT}}$ ). In addition, for deriving equation (3.6) we have adopted an adiabatic index,  $\gamma = 5/3$ , mean particle mass,  $\mu = 1.23 m_{\text{H}}$ . After eliminating  $N_{\text{H}}$  between equation (3.5) and equation (3.6), the photoionization rate can be written as a function of the hydrogen number density

$$\Gamma_{\star} \sim 1.3 \times 10^{-13} \text{ s}^{-1} \left( \frac{n_{\text{H}}}{1 \text{ cm}^{-3}} \right)^{0.2} \left( \frac{f_{\text{g}}}{f_{\text{th}}} \right)^{0.2} T_4^{0.2}. \quad (3.7)$$

Based on equation (3.7), the photoionization rate produced by star formation is only weakly sensitive to the temperature, total gas fraction,  $f_{\text{g}}$ , and the fraction of the pressure that is thermal,  $f_{\text{th}}$ . The photoionization rate in equation (3.7) is also very weakly dependent on the gas density. As we will discuss in §3.4.1, our simulations show the same trend and are in excellent agreement with this analytic estimate. However, one should note that due to the simplified assumptions we adopted to derive equation (3.7), it provides only an order-of-magnitude estimate for the typical photoionization rate that is produced by young stars in the ISM, on kilo-parsec scales. This relation does not capture the inhomogeneity of the ISM that may result in large fluctuations in the radiation field on smaller scales. We note that the relation we found here between the gas density and stellar photoionization rate is a direct consequence of the underlying star formation law and is independent of the total amount of star formation in a galaxy.

### 3.3 Simulation techniques

In this section we describe different parts of our simulations. We start with discussing the details of the hydrodynamical simulations that are post-processed with RT calculations. Then we explain our RT that includes the radiation from local stellar radiation, the UVB and recombination radiation.

#### 3.3.1 Hydrodynamical simulations

We use a set of cosmological simulations that are performed using a modified and extended version of the smoothed particle hydrodynamics (SPH) code GADGET-3 (last described in Springel, 2005). The physical processes that are included in the simulations are identical to what has been used in the reference simulation of the Overwhelmingly Large Simulations (OWLS) described in Schaye et al. (2010). Briefly, we use a subgrid pressure-dependent star formation prescription of Schaye & Dalla Vecchia (2008) which reproduces the observed



Kennicutt-Schmidt law. We use the chemodynamics model of Wiersma et al. (2009b) which follows the abundances of eleven elements assuming a Chabrier (2003) IMF. These abundances are used for calculating radiative cooling/heating rates, element-by-element and in the presence of the uniform cosmic microwave background and the Haardt & Madau (2001) UVB model (Wiersma et al., 2009a). We note that the Haardt & Madau (2001) UVB model has been shown to be consistent with observations of H $\alpha$  (Altay et al., 2011; Rahmati et al., 2013) and metal absorption lines (Aguirre et al., 2008). We model galactic winds driven by star formation using a kinetic feedback recipe that assumes 40% of the kinetic energy generated by Type II SNe is injected as outflows with initial velocity of 600 km s $^{-1}$  and with a mass loading parameter  $\eta = 2$  (Dalla Vecchia & Schaye, 2008).

We adopt cosmological parameters consistent with the most recent WMAP year 7 results:  $\{\Omega_m = 0.272, \Omega_b = 0.0455, \Omega_\Lambda = 0.728, \sigma_8 = 0.81, n_s = 0.967, h = 0.704\}$  (Komatsu et al., 2011). Our reference simulation has a periodic box of  $L = 6.25$  comoving  $h^{-1}$ Mpc and contains  $128^3$  dark matter particles with mass  $6.3 \times 10^6 h^{-1}M_\odot$  and an equal number of baryons with initial mass  $1.4 \times 10^6 h^{-1}M_\odot$ . The Plummer equivalent gravitational softening length is set to  $\epsilon_{\text{com}} = 1.95 h^{-1}$ kpc and is limited to a minimum physical scale of  $\epsilon_{\text{prop}} = 0.5 h^{-1}$ kpc. We also use a simulation with a box size identical to our reference simulation but with 8 (2) times better mass (spatial) resolution to assess the effect of resolution on our findings (see Appendix B2).

In our hydrodynamical simulations, ISM gas particles (which all have densities  $n_H > 0.1 \text{ cm}^{-3}$ ) follow a polytropic equation of state that defines their temperatures. These temperatures are not physical and only measure the imposed pressure (Schaye & Dalla Vecchia, 2008). Therefore, when calculating recombination and collisional ionization rates, we set the temperature of ISM particles to  $T_{\text{ISM}} = 10^4$  K which is the typical temperature of the warm-neutral phase of the ISM. Furthermore, we simplify our RT calculations by assuming that helium and hydrogen absorb the same amount of ionizing radiation per unit mass and by ignoring dust absorption and the possibility that some hydrogen may be molecular (see Rahmati et al., 2013 for more discussion).

### 3.3.2 Radiative transfer

For the RT calculations we use TRAPHIC (see Pawlik & Schaye, 2008, 2011; Rahmati et al., 2013). TRAPHIC is an explicitly photon-conserving RT method designed to exploit the full spatial resolution of SPH simulations by transporting radiation directly on the irregular distribution of SPH particles.

The RT calculation in TRAPHIC starts with source particles emitting photon packets to their neighbors. This is done using a set of  $N_{\text{EC}}$  tessellating emission cones, each subtending a solid angle of  $4\pi/N_{\text{EC}}$ . The propagation directions of the photon packets are initially parallel to the central axes of the emission cones. In order to improve the angular sampling of the RT, the orientations of these

emission cones are randomly rotated between emission time steps. We adopt  $N_{\text{EC}} = 8$  for computational efficiency. However, we note that our results are insensitive to the precise value of  $N_{\text{EC}}$ , thanks to the random rotations.

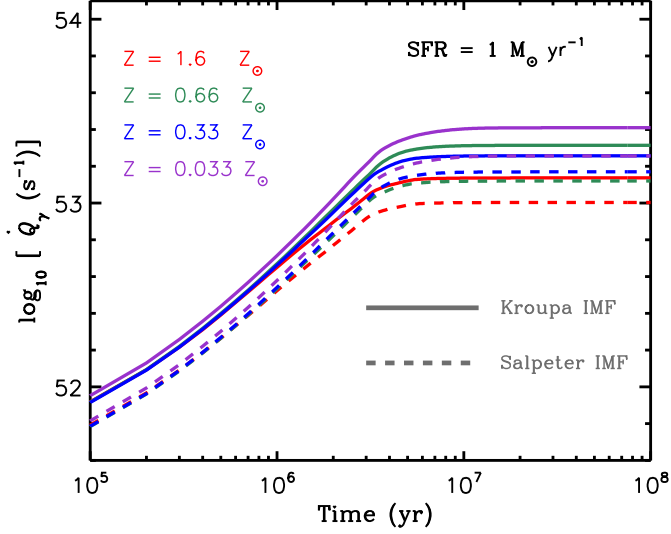
After emission, photon packets travel along their propagation direction from one SPH particle to its neighbors. Only neighbors that are inside transmission cones can receive photons. Transmission cones are defined as regular cones with opening solid angle  $4\pi/N_{\text{TC}}$ , and are centered on the propagation direction. The parameter  $N_{\text{TC}}$  sets the angular resolution of the RT and we adopt  $N_{\text{TC}} = 64$  which produces converged results (see Appendix C1 of [Rahmati et al., 2013](#)). To guarantee the independence of the RT from the distribution of SPH particles, additional virtual particles (ViP) are introduced. This is done wherever a transmission cone contains no neighboring SPH particle. ViPs do not affect the underlying SPH simulation and are deleted after the photon packets are transferred.

Furthermore, arriving photon packets are merged into a discrete number of reception cones,  $N_{\text{RC}} = 8$ . This makes the computational cost of RT calculations with TRAPHIC independent of the number of radiation sources. This feature is particularly important for the purpose of the present work as it enables the RT calculations in cosmological density fields with large numbers of sources.

Reception cones are also used for emission from gas particles (e.g., ionizing radiation from star-forming gas particles, recombination radiation). This reduces computational expenses while yielding accurate results. Photon packets are isotropically emitted into reception cones which are already in place at each SPH particle. This obviates the need for constructing any additional emission cone tessellations. Using this recipe, the emission of radiation by star-forming gas particles is identical to that of the emission of diffuse radiation by recombining gas particles, and is described in more detail in Raicevic et al. (in prep.).

Photon packets emitted by the three main radiation sources we consider in our study (i.e., UVB, RR and stellar radiation) are channeled in separate frequency bins and are not merged with each other. This enables us to compute the contribution of each component to the total photoionization rate. The total amount of the absorbed radiation, summed over all frequency bins, determines the ionization state of the absorbing SPH particles. The time-dependent differential equations that control the evolution of the ionization states of different species (e.g., H, He) are solved using a sub-cycling scheme that allows us to choose the RT time step independent of the photoionization and recombination times.

In our RT calculations we use a set of numerical parameters identical to that used in [Rahmati et al. \(2013\)](#). These parameters produce converged results. In addition to the parameters mentioned above, we use 48 neighbors for SPH particles, 5 neighbors for ViPs and an RT time step of  $\Delta t = 1 \text{ Myr} \left( \frac{4}{1+z} \right)$ . Ionizing photons are propagated at the speed of light, inside the simulation box with absorbing boundaries, until the equilibrium solution for the hydro-



**Figure 3.1:** The number of hydrogen ionizing photons produced for a constant star formation rate of  $1 M_{\odot} \text{ yr}^{-1}$  as a function of time since the onset of star formation. These results are calculated using STARBURST99. Red, green, blue and purple curves indicate metallicities of  $Z/Z_{\odot} = 1.6$ ,  $6.6 \times 10^{-1}$ ,  $3.3 \times 10^{-1}$ , &  $3.3 \times 10^{-2}$ , respectively. Solid and dashed curves show results for the Kroupa and Salpeter IMF, respectively. All the curves with different metallicities and IMFs converge to similar equilibrium values at  $\dot{Q}_{\gamma} \sim 2 \times 10^{53}$  photons per second, after about  $\sim 5 - 10$  Myr.

gen neutral fractions is reached. To facilitate this process, the RT calculation is starting from an initial neutral state, except for the gas at low densities (i.e.,  $n_{\text{gas}} < 1 \times 10^{-3} \text{ cm}^{-3}$ ) and high temperatures (i.e.,  $T > 10^5 \text{ K}$ ) which is assumed to be in equilibrium with the UVB photoionization rate and its collisional ionization rate. Typically, the average neutral fraction in the simulation box does not evolve after 2-3 light-crossing times (the light-crossing time for the  $L_{\text{box}} = 6.25$  comoving  $h^{-1} \text{ Mpc}$  is  $\approx 7.2 \text{ Myr}$  at  $z = 3$ ).

We continue this section by briefly describing the implementation of the UVB, diffuse recombination radiation and stellar radiation.

### 3.3.3 Ionizing background radiation and diffuse recombination radiation

In principle, local sources of ionization inside the simulation box should be able to generate the UVB. However, the box size of our simulations is smaller than the mean free path of ionizing photons which makes the simulated volume too small to generate the observed UVB intensity (see §3.4.3). In addition, a considerable fraction of the UVB at  $z \lesssim 3$  is produced by quasars which are not included in

our simulations. Therefore, we impose an additional UVB in our simulation box.

The implementation of the UVB is identical to that of [Rahmati et al. \(2013\)](#). The ionizing background radiation is simulated as plain-parallel radiation entering the simulation box from its sides and the injection rate of the UVB ionizing photons is normalized to the desired photoionization rate in the absence of any absorption (i.e., the so-called optically thin limit).

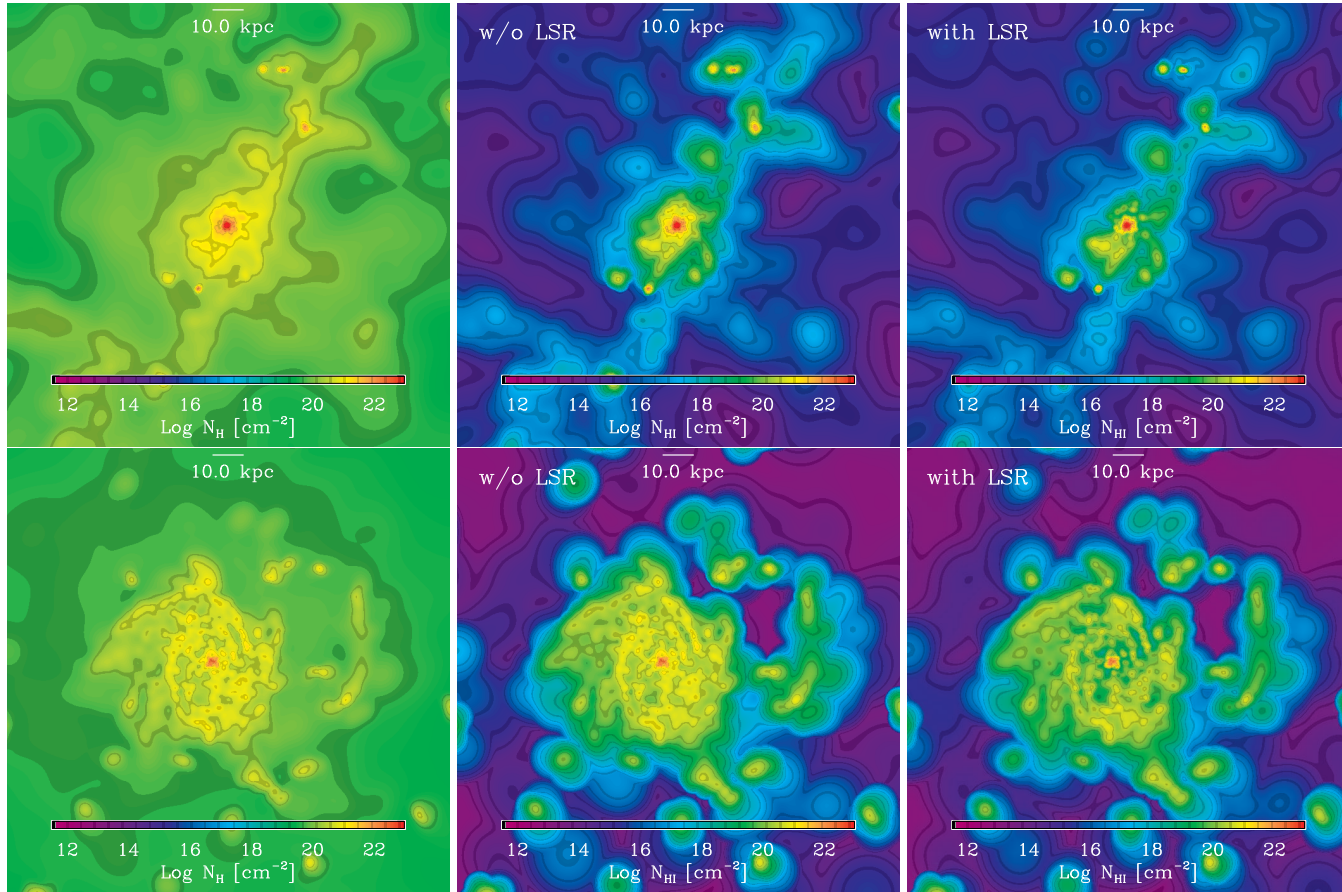
We set the effective photoionization rate and spectral shape of the UVB radiation at different redshifts based on the UVB model of [Haardt & Madau \(2001\)](#) for quasars and galaxies. The same UVB model has been used to calculate heating/cooling in our hydrodynamical simulations and has been shown to be consistent with observations of H I ([Altay et al., 2011; Rahmati et al., 2013](#)) and metal absorption lines ([Aguirre et al., 2008](#)) at  $z \sim 3$ . We adopt the gray approximation instead of an explicit multifrequency treatment for the UVB radiation. Our tests show that a multifrequency treatment of the UVB radiation does not significantly change the resulting hydrogen neutral fractions (see Appendix D1 of [Rahmati et al., 2013](#)).

In addition, hydrogen recombination radiation (RR) is simulated by making all SPH particles isotropic radiation sources with emissivities based on their recombination rates. We do not account for the redshifting of the recombination photons and assume that they are monochromatic with energy 13.6 eV (see [Raicevic et al. in prep.](#)).

### 3.3.4 Stellar ionizing radiation

The ionizing photon production rate of star-forming galaxies is dominated by young and massive stars which have relatively short life times. In cosmological simulations with limited mass resolutions, the spatial distribution of newly formed stellar particles (e.g., with ages less than a few tens of Myr) may sample the locally imposed star formation rates relatively poorly. As we show in §3.4.2, the distribution of star formation is better sampled by star-forming SPH particles and this is particularly important for low and intermediate halo masses. For this reason, we use star-forming particles as sources of stellar ionizing radiation.

For a constant star formation rate, the production rate of ionizing photons reaches an equilibrium value within  $\sim 5 - 10$  Myr (i.e., the typical life-time of massive stars). This equilibrium photon production rate per unit star formation rate can be calculated using stellar population synthesis models. We used STARBURST99 ([Leitherer et al., 1999](#)) to calculate this emissivity for the [Kroupa \(2001\)](#) and [Salpeter \(1955\)](#) IMF and a metallicity consistent with the typical metallicities of star-forming particles in our simulations (the median metallicity of star-forming particles in our simulations are within the range  $10^{-1} \lesssim Z/Z_{\odot} \lesssim 1$  at redshifts 0 - 5). We found that for a constant star formation rate of  $1 M_{\odot} \text{ yr}^{-1}$  and after  $\sim 10$  Myr, the photon production rate of ionizing photons converges to  $\dot{Q}_{\gamma} \sim 2 \times 10^{53} \text{ s}^{-1}$ , which is the value we used in equation (3.2). As shown in



**Figure 3.2:** The H I column density distribution of a Milky Way-like galaxy in the reference simulation at  $z = 3$  (top) and  $z = 0$  (bottom). The left panels show the total hydrogen column density distribution. The middle panels show the H I column densities in the presence of collisional ionization, photoionization from the UVB and recombination radiation. In the right panel, the H I distribution is shown after photoionization from local stellar radiation is added to the other sources of ionization. The images are  $100 h^{-1} \text{kpc}$  (proper) on a side. Comparison between the right and middle panels shows that local stellar radiation substantially changes the H I column density distribution at  $N_{\text{HI}} \sim 10^{21} \text{ cm}^{-2}$ .



Figure 3.1, the ionizing photon production rate varies only by  $\lesssim \pm 0.2$  dex if the metallicity changes by  $\sim \pm 1$  dex. Also, reasonable variations in the IMF (e.g., the Chabrier IMF) do not significantly change our adopted value for the photon production rate. For the spectral shape of the stellar radiation, we adopt a blackbody spectrum with temperature  $T_{\text{bb}} = 5 \times 10^4$  K which is consistent with the spectrum of massive young stars.

## 3.4 Results and discussion

In this section, we report our findings based on RT simulations that include the UVB, recombination radiation and local stellar radiation. The RT calculations are performed by post-processing a hydrodynamical simulation with  $128^3$  SPH particles in a  $6.25$  comoving  $h^{-1}\text{Mpc}$  box and at redshifts  $z = 0 - 5$ . In §3.4.1, we compare different photoionizing components and assess the impact of local stellar radiation on the H I distribution. In §3.4.2 we illustrate the importance of using star-forming SPH particles instead of using young stellar particles as ionizing sources. In §3.4.3 we show that in our simulations, the predicted intensity of stellar radiation in the IGM is consistent with the observed intensity of the UVB and calculate the implied average escape fraction. In §3.4.4 we show that assumptions about the unresolved properties of the ISM are very important and we discuss the impact of local stellar radiation on the H I column density distribution.

### 3.4.1 The role of local stellar radiation in hydrogen ionization

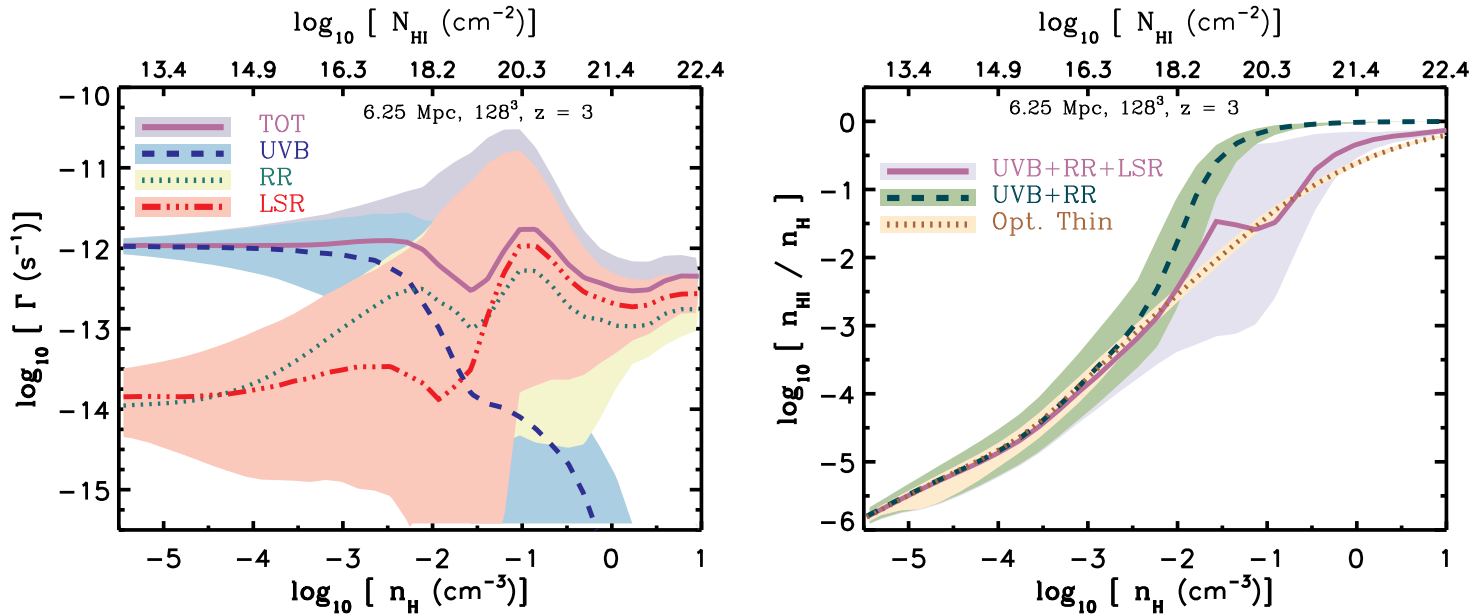
As we showed in Rahmati et al. (2013), the UVB radiation and collisional ionization are the dominant sources of ionization at low densities where the gas is shock heated and optically thin. However, self-shielding prevents the UVB radiation from penetrating high-density gas. In addition, much of the ISM has temperatures that are too low for collisional ionization to be efficient. Therefore, because UVB photoionization and collisional ionization are both inefficient in those regions, the ionizing radiation from young stars becomes the primary source of ionization.

Figure 3.2 shows the distribution of neutral hydrogen in and around a galaxy in our reference simulation at redshifts  $z = 3$  (top row) and  $0$  (bottom row) and illustrates that the distribution of H I in galaxies may change significantly as a result of local stellar radiation. The total mass of the halo hosting this galaxy at  $z = 3$  and  $0$  is  $M_{200} = 2 \times 10^{11}M_{\odot}$  and  $1.1 \times 10^{12}M_{\odot}$ , respectively. Comparing the panels in the left and middle columns, we see that while collisional ionization and photoionization by the UVB ionize the low-density gas around the galaxy, they do not change the high H I column densities in the inner regions. Comparing the yellow regions in the right and middle panels of Figure 3.2 demonstrates that local stellar radiation significantly changes the distribu-

tion of the HI at  $N_{\text{HI}} \sim 10^{21} \text{ cm}^{-2}$ . In the left panel of Figure 3.3 the contributions of the photoionization rates from different radiation components are plotted against the total hydrogen number density at  $z = 3$ . The purple solid line shows the total photoionization rate. The blue dashed, green dotted and red dot-dashed lines show respectively the contribution of the UVB, recombination radiation (RR) and local stellar radiation (LSR). The resulting hydrogen neutral fraction as a function of gas density is shown by the purple solid curve in the right panel of Figure 3.3 which illustrates the significant impact of LSR at high densities. We note that there is a sharp feature in the hydrogen neutral fraction at  $10^{-1} < n_{\text{H}} < 10^{-2} \text{ cm}^{-3}$  in the presence of local stellar radiation. This feature corresponds to a sharp drop-off in the photoionization rate from local stellar radiation at the same densities (see the red dot-dashed curve in the left panel of Figure 3.3). The left panel of Figure 3.4 shows photoionization rate profiles in spherical shells around haloes with  $10.5 < \log_{10} M_{200} < 11 M_{\odot}$  at  $z = 3$ . The right panel of Figure 3.4 shows the resulting neutral hydrogen fraction profile (purple solid curve) which is compared with the simulation that does not include local stellar radiation (green dashed curve). We note that the trends in the photoionization rate and hydrogen neutral fraction profiles do not strongly depend on the chosen mass bin as long as  $\log_{10} M_{200} \gtrsim 9.5 M_{\odot}$ .

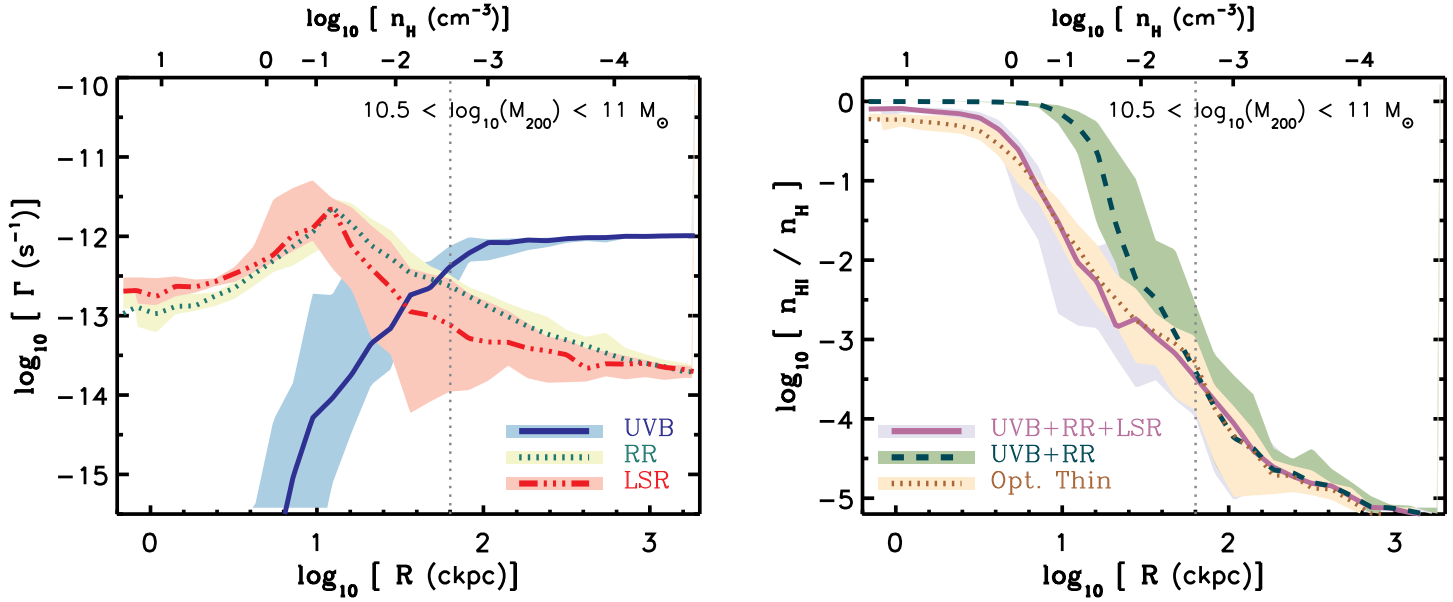
As Figures 3.3 and 3.4 show, at low densities (i.e., at large distances from the centers of the halos) the gas is highly ionized by a combination of the UVB and collisional ionization and this optically thin gas does not absorb a significant fraction of the ambient ionizing radiation. Consequently, at  $n_{\text{H}} \lesssim 10^{-4} \text{ cm}^{-3}$ , which corresponds to typical distances  $R \gtrsim 1$  comoving Mpc from the centers of the haloes, the UVB photoionization rate does not change with density. By increasing the density, or equivalently by decreasing the distance to the center of the haloes, the optical depth increases and eventually the gas becomes self-shielded against the UVB. This causes a sharp drop in the UVB photoionization rate at the self-shielding density (see the blue dashed curves in the left panels of Figures 3.3 and 3.4). As shown in Rahmati et al. (2013), the UVB photoionization rate shows a similar behavior in the absence of local stellar radiation. However, local stellar radiation increases the photoionization rate at high and intermediate densities. This decreases the hydrogen neutral fractions around the self-shielding densities, allowing the UVB radiation to penetrate to higher densities. Consequently, the local stellar radiation increases the effective self-shielding threshold against the UVB radiation slightly (by  $\sim 0.1$  dex) compared to the simulation without the radiation from local stars (not shown).

As shown by the red dot-dashed curve in the left panel of Figure 3.3, at densities  $0.1 \lesssim n_{\text{H}} \lesssim 1 \text{ cm}^{-3}$  the median photoionization rate due to local stellar radiation increases with decreasing density. The main reason for this is the superposition of radiation from multiple sources (i.e., star-forming SPH particles) as the mean free path of ionizing photons increases with decreasing density (see Figure 3.11). This is also seen in the left panel of Figure 3.4 as increasing stellar photoionization rate with increasing the distance from the center of halos



**Figure 3.3:** *Left:* Photoionization rate as a function of density due to different radiation components in the reference simulation at  $z = 3$ . The purple solid curve shows the total photoionization rate. The blue dashed, green dotted and red dot-dashed curves show respectively the photoionization rates due to the UVB, diffuse recombination radiation (RR) and local stellar radiation (LSR). *Right:* The hydrogen neutral fraction as a function of density for the same simulation is shown with the purple solid line. For comparison, the results for a simulation without local stellar radiation (green dashed curve) and a simulation with the UVB radiation in the optically thin limit (i.e., no absorption; brown dotted curve) are also shown. The curves show the medians and the shaded areas around them indicate the 15% – 85% percentiles. HI column densities corresponding to each density in the presence of all ionization sources are shown along the top x-axis. The photoionization due to local stellar radiation exceeds the UVB photoionization rate at high densities and compensates the effect of self-shielding. This produces a hydrogen neutral fraction profile that is very similar to what is expected from the UVB in the optically thin limit.





**Figure 3.4:** *Left:* Median photoionization rate profiles (comoving) due to different radiation components for haloes with  $10.5 < \log_{10} M_{200} < 11 M_{\odot}$  at  $z = 3$ . Blue solid, green dotted and red dot-dashed curves show the photoionization rates of the UVB, recombination radiation (RR) and local stellar radiation (LSR). *Right:* Median hydrogen neutral fraction profiles for the same haloes with local stellar radiation (purple solid curve), without local stellar radiation (green dashed curve) and with only the UVB radiation in the optically thin limit (dotted brown). In both panels, the vertical dotted line indicates the median  $R_{200}$  radius of the haloes in the chosen mass bin. The shaded areas around the medians indicate the 15% – 85% percentiles. The *top axis* in each panel shows the median density at a given comoving distance from the center of the haloes. The photoionization due to local stellar radiation exceeds the UVB photoionization rate close to galaxies and compensates the effect of self-shielding. This produces a hydrogen neutral fraction profile that is very similar to what is expected from the UVB in the optically thin limit.

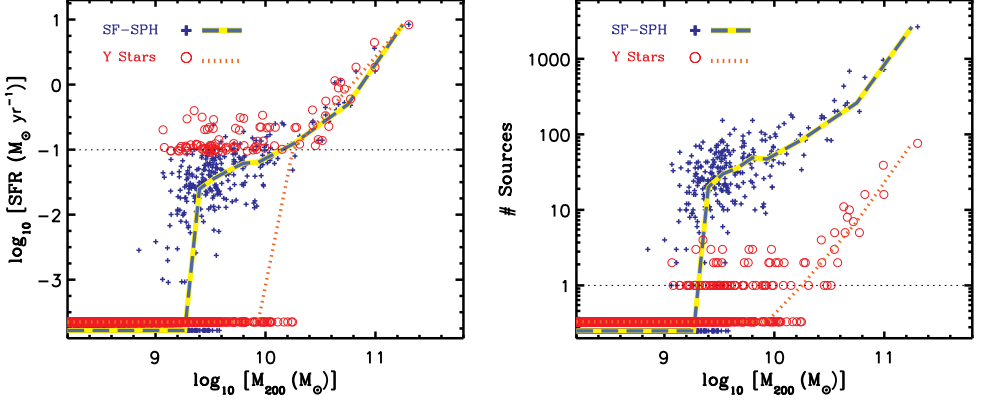
for  $R < 1$  comoving kpc. On the other hand, at densities lower than the star formation threshold (i.e.,  $n_{\text{H}} < 0.1 \text{ cm}^{-3}$ ), the gas is typically at larger distances from the star-forming regions. Therefore, the photoionization rate of local stellar radiation drops rapidly with decreasing density. The star formation rate density averaged on larger and larger scales becomes increasingly more uniform. This causes the photoionization from galaxies that are emitting ionizing radiation to produce a more uniform photoionization rate at the lowest densities. Note that at the highest densities, the photoionization rate from local stellar radiation agrees well with the analytic estimate presented in §3.2.

In the absence of local stellar radiation, recombination radiation photoionization rate peaks at around the self-shielding density (Rahmati et al., 2013). The reason for this is that the production rate of ionizing photons by recombination radiation is proportional to the density of the ionized gas. At densities below the self-shielding threshold, hydrogen is highly ionized and recombination rate is proportional to the total hydrogen density. At much higher densities, the gas is nearly neutral and little recombination radiation is generated. As the left panel in Figure 3.3 shows, at high densities the situation changes dramatically if we include local stellar radiation. Since the gas at high densities (e.g.,  $n_{\text{H}} \gg 10^{-2} \text{ cm}^{-3}$ ) is optically thick, recombination radiation ionizing photons produced at these densities are absorbed locally. In equilibrium, recombination radiation photoionization rate at densities above the self-shielding is therefore a constant fraction of the total ionization rate<sup>1</sup>. Below the self-shielding density, the recombination radiation photoionization rate decreases with decreasing density and asymptotes to a background rate. However, in reality recombination photons cannot travel to large cosmological distances without being redshifted to frequencies below the Lyman edge. Therefore, our neglect of the cosmological redshifting of recombination radiation leads us to overestimate the photoionization rate due to recombination radiation on large scales. On the other hand, because of the small size of our simulation box, the total photoionization rate that is produced by recombination radiation remains negligible compared to the UVB photoionization rate and neglecting the redshifting of recombination radiation is not expected to affect our results.

The purple solid curve in the right panel of Figure 3.3 shows the hydrogen neutral fractions in the presence of the UVB, recombination radiation and local stellar radiation. For comparison, the hydrogen neutral fractions in the absence of local stellar radiation, and for the optically thin gas that is photoionized only by the UVB, are also shown (with the green dashed and brown dotted curves respectively). Hydrogen at densities  $n_{\text{H}} \gtrsim 10^{-1} \text{ cm}^{-3}$  is self-shielded and mostly neutral if the UVB and recombination radiation are the only sources of photoion-

---

<sup>1</sup>This can be explained noting that in equilibrium, recombination and ionization rates are equal. Depending on temperature,  $\approx 40\%$  of recombination photons are ionizing photons which are absorbed on the spot in optically thick gas (e.g., Osterbrock & Ferland, 2006). Therefore, the photoionization rate produced by recombination radiation is also  $\approx 40\%$  of the total ionization rate in dense and optically thick gas with  $T \sim 10^4 \text{ K}$ .



**Figure 3.5:** Star-formation activity of haloes in the reference simulation (6.25 Mpc ,  $128^3$ ) at  $z = 3$ . In the *left panel* the blue crosses show the total instantaneous star formation rate for a given halo computed using the gas particles and the red circles show the star formation rate calculated by dividing the total stellar mass formed during the last 20 Myr by 20 Myr. Haloes with zero star formation rates are shown in the bottom of the plot. For massive haloes the two measures agree but as a result of limited mass resolution and the stochastic nature of the star formation algorithm, they start to differ substantially for haloes with  $M_{200} \lesssim 10^{10} M_{\odot}$ . A large fraction of low-mass haloes does not contain any young stellar particles and the median star formation rate calculated using young stellar particles drops to zero. The *right panel* shows the number of ionizing sources for different halos in the same simulation. While the blue crosses show the number of star-forming gas particles in each halo, red circles indicate the number of young stellar particles (i.e., younger than 20 Myr). In both panels the horizontal dotted lines correspond to a single stellar particle and the blue dashed and red dotted curves show respectively the medians for star-forming gas and young stellar particles.

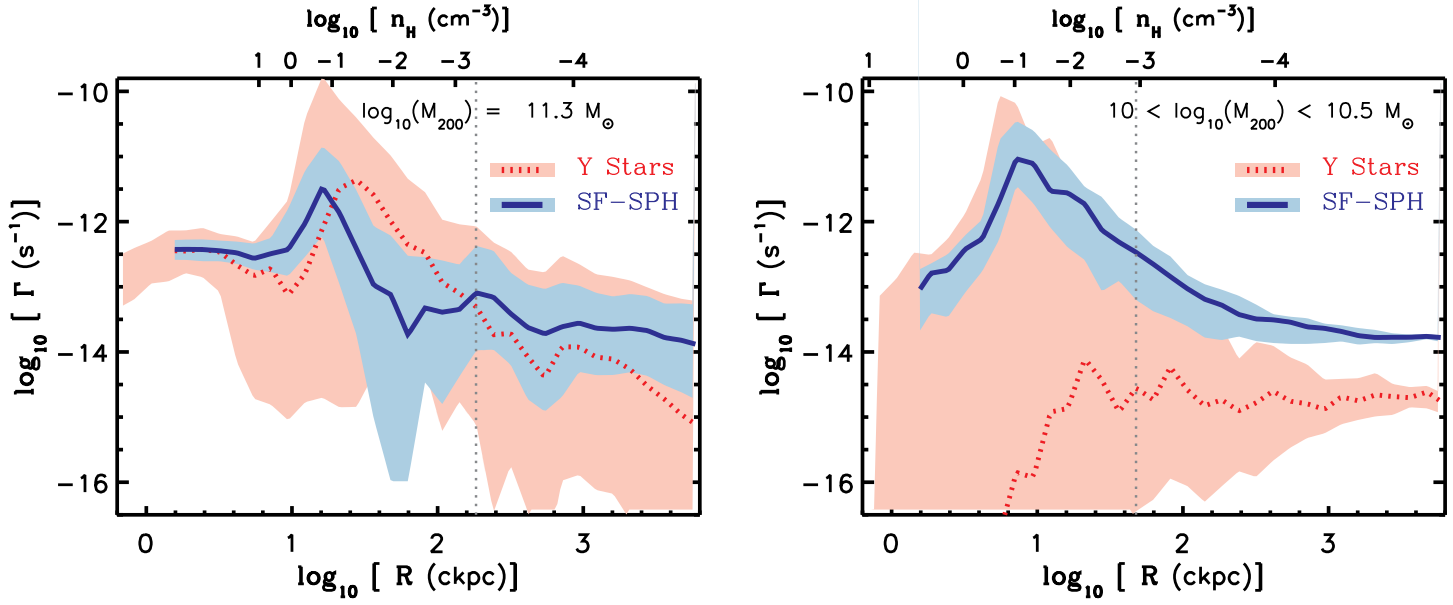
ization. However, local stellar radiation significantly ionizes the gas at intermediate and high densities. As mentioned in §3.2, the typical photoionization rate that is produced by star-forming gas is  $\Gamma_{\text{SF}} \sim 10^{-13} \text{ s}^{-1}$ , which is comparable to the photoionization rate of the unattenuated UVB at  $z \sim 0$ . Consequently, the hydrogen neutral fractions in the presence of local stellar radiation are much closer to the optically thin case. It is also worth noting that the scatter around the median hydrogen neutral fractions is largest for intermediate densities (i.e.,  $10^{-3} \lesssim n_{\text{H}} \lesssim 1 \text{ cm}^{-3}$ ). This is closely related to the large scatter in the photoionization rate produced by local stellar radiation (see the left panel in Figure 3.3). At these densities, the large scatter in the distances to the nearest sources and RT effects like shadowing produce a large range of photoionization rates. For the gas at very high and very low densities on the other hand, the scatter becomes smaller because the relative distribution of sources with respect to the absorbing gas becomes more uniform. The trends discussed so far are qualitatively similar at redshifts other than  $z = 3$ : Although the star formation rates

evolve at high densities, the photoionization rate due to local stars is set by the underlying star formation law (see §3.2) which does not change with time. The peak of the stellar photoionization rate at  $n_{\text{H}} \sim 10^{-1} \text{ cm}^{-3}$ , produced by the superposition of multiple sources, exists at all redshifts but it moves to slightly higher densities at lower redshifts. At densities immediately below the adopted star formation threshold (i.e.,  $n_{\text{H}} \lesssim 0.1 \text{ cm}^{-3}$ ), the photoionization rate produced by local sources drops rapidly with decreasing density. The distribution of star formation in the simulation box is almost uniform on large scales. Therefore, at the lowest densities, local stellar radiation produces a photoionization rate which is not changing strongly with density. However, at low redshifts (e.g.,  $z \lesssim 1$ ), the star formation density decreases significantly and becomes highly non-uniform on the scales probed by our small simulation box. Consequently, the resulting photoionization rate due to young stars does not converge to a constant value at low densities at these redshifts. For the same reason, at low densities the scatter around the median stellar photoionization rate increases with decreasing redshift (not shown). As we will discuss in §3.4.3, if we correct for the small size of our simulation box, the asymptotic photoionization rate due to stellar radiation that has reached the IGM, is consistent with the intensity of the observed UVB at the same redshift.

### 3.4.2 Star-forming particles versus stellar particles

In cosmological simulations, the integrated instantaneous star formation rate of the simulation box closely matches the total amount of mass converted into stellar particles. However, due to limited resolution, the spatial distribution of young stellar particles (e.g., those with ages  $\sim 10$  Myr) may not sample the spatial distribution of star formation in individual haloes very well.

This issue is illustrated in the left panel of Figure 3.5, where the blue crosses show the total instantaneous star formation rates inside galaxies in our reference simulation at  $z = 3$ . These star formation rates are calculated by the summation of the star formation rates of all SPH particles in a given galaxy. The star formation rates can also be estimated by measuring the average rate by which stellar particles are formed in a given galaxy over some small time interval. The star formation rate calculated using this method (i.e., measuring the rate by which stellar particles are formed during the last 20 Myr), is indicated in the left panel of Figure 3.5 by the red circles. For massive galaxies, with high star formation rates, the formation rate of stellar particles agrees reasonably well with the star formation rates computed from the gas distribution. However, most haloes with  $\lesssim 10^3$  SPH particles (i.e.,  $M_{200} \lesssim 10^{10} M_{\odot}$ ) do not contain any young stellar particles despite having non-zero instantaneous star formation rates. Consequently, for the few low-mass galaxies that by chance contain one or more young stellar particles, the implied star formation rates are much higher than the instantaneous rate that corresponds to the gas distribution. Moreover, as the right panel of Figure 3.5 shows, the number of star-forming particles in a given simulated



**Figure 3.6:** Median photoionization rate profiles (comoving) for local stellar radiation emitted by star-forming SPH particles (SF-SPH; blue solid curves) and young ( $< 20$  Myr) stellar particles (Y Stars; red dotted curves) at  $z = 3$ . The *left panel* shows the photoionization rate profiles for a halo with  $M_{200} = 2 \times 10^{11} M_{\odot}$  while the *right panel* shows the same profiles for haloes with  $10 < \log_{10} M_{200} < 10.5 M_{\odot}$ . In both panels the vertical dotted line indicates the median  $R_{200}$  radius of the halos in the illustrated mass bin. The shaded areas around the medians indicate the 15% – 85% percentiles. The *top axis* in each panel shows the median density at a given comoving distance from the centers of the haloes. While the photoionization rate profiles produced by star-forming gas particles and young stellar particles are similar for massive haloes (left), they are substantially different for haloes with  $M_{200} \lesssim 10^{10} M_{\odot}$  (right).

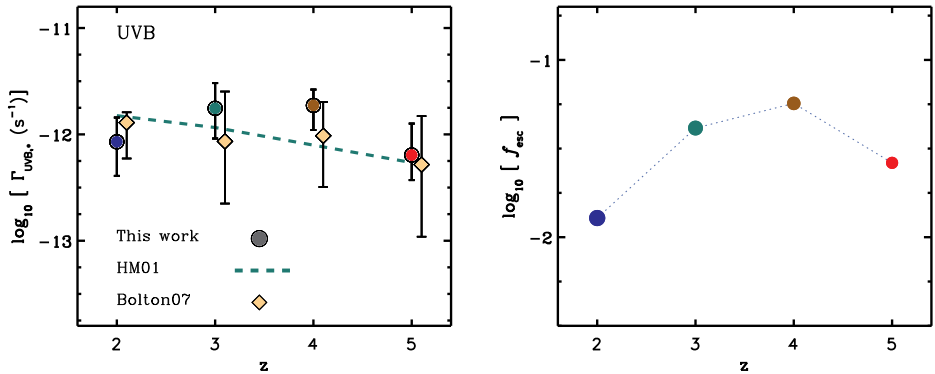
galaxy is  $\sim 10^2$  times larger than that of young stellar particles. This ratio could be understood by noting that the observed star formation law implies a gas consumption time scale in the ISM that is  $\sim 10^2$  times longer than the life times of massive stars. Hence, if star formation is implemented by the stochastic conversion of gas particles into stellar particles, as is the case here, the number of star-forming (i.e., ISM) particles is expected to be  $\sim 10^2$  times larger than the number of young stellar particles. This ratio will be somewhat smaller in starbursts or if gas particles are allowed to spawn multiple stellar particles, but under realistic conditions, star formation will be sampled substantially better by gas particles than by young stellar particles.

As a result of the above mentioned sampling effects, using stellar particles as ionizing sources (as was done in all previous work, e.g., Nagamine et al., 2010; Fumagalli et al., 2011; Yajima et al., 2012) would underestimate the impact of local stellar radiation on the H I distribution for a large fraction of simulated galaxies. We therefore use star-forming SPH particles instead of young stellar particles as local sources of radiation.

Figure 3.6 illustrates the difference between the photoionization rate profiles produced by star-forming SPH particles (blue solid curves) and young stellar particles (red dotted curves) at  $z = 3$ . The left panel of Figure 3.6 shows this for a halo with  $M_{200} = 2 \times 10^{11} M_{\odot}$  (see the top panels of Figure 3.2), while the right panel illustrates the results for haloes with  $10 < \log_{10} M_{200} < 10.5 M_{\odot}$ . For this comparison we imposed the same total number of emitted photons in the simulation box in both cases. This was done by setting the photon production rates of individual gas particles proportional to their star formation rates (see equation 3.2) and setting the photon production rate of individual young stellar particles proportional to their masses.

For the massive halo, simulating local stellar radiation using SPH particles results in photoionization rate profiles similar to those produced by using stellar particles. This similarity is mainly due to the agreement between the total star formation rate and the rate by which gas is converted into young stellar particles (see the left panel of Figure 3.5). However, this galaxy has  $\sim 50$  times fewer young stellar particles than star-forming SPH particles (see the right panel of Figure 3.5) and as the shaded areas in the left panel of Figure 3.6 show, the scatter in the photoionization rate profile is much larger when stellar particles are used.

For haloes with slightly lower masses, shown in the right panel of Figure 3.6, the rates by which gas is converted into young stellar particles are similar to the total star formation rates (see the left panel of Figure 3.5). Despite this agreement, the median photoionization rates for these two cases differ dramatically, being much higher when star-forming gas particles are used as sources. We conclude that using stellar particles as sources results in the underestimation of the ionization impact of local stellar radiation on most of the galaxies in the simulation. The intensity of the UVB produced by the simulation that uses young stellar particles as sources does not agree with the observed UVB intens-



**Figure 3.7:** *Left:* The simulated UVB photoionization rate produced by stellar radiation in our reference simulation is shown with filled circles for different redshifts. The observed mean free paths of ionizing photons have been used to correct for the small size of the simulation box (see text). The error bars represent the  $1 - \sigma$  errors in these mean free paths. The dashed curve shows the Haardt & Madau (2001) UVB photoionization rates that have been used in our simulations as the UVB. The observational measurement of the UVB from the Ly $\alpha$  effective opacity by Bolton & Haehnelt (2007) is shown using orange diamonds. *Right:* the average escape fraction of stellar ionizing radiation into the IGM calculated based on equation 3.13. Our simulation reproduces the observed UVB photoionization rates between  $z = 2$  and 5. The implied average escape fractions are  $10^{-2} < f_{\text{esc}} < 10^{-1}$  between  $z = 2$  and 5 and they decrease with decreasing redshift below  $z = 4$

ity, while using star-forming particles resolves this issue, after correcting for the box size (see §3.4.3).

Although in this section we have shown that using star-forming particles as sources of ionizing radiation helps to resolve the above mentioned sampling issues in post-processing RT simulations, we note that doing the same may not be a good solution for simulations with sufficient resolution to model the cold, interstellar gas phase. Such simulations can capture the effects of the relative motions of stars and gas, e.g., the effect of stars moving out of, or destroying, their parent molecular clouds.

### 3.4.3 Stellar ionizing radiation, its escape fraction and the buildup of the UVB

Observations show that the luminosity function of bright quasars drops sharply at redshifts  $z \gtrsim 2$  (e.g., Hopkins et al., 2007) and models for the cosmic ionizing background indicate that star-forming galaxies dominate the production of hydrogen ionizing photons at  $z \gtrsim 3$  (e.g., Haehnelt et al., 2001; Bolton et al., 2005; Faucher-Giguère et al., 2008a). In the simulations, it should therefore be possible to build up the background radiation from the radiation produced by

**Table 3.1:** The comoving mean free path of hydrogen Lyman-limit photons,  $\lambda_{\text{mfp}}$ , at different redshifts. From left to right, columns respectively show redshift,  $\lambda_{\text{mfp}}$  in comoving Mpc (cMpc) and the references from which the mean free path values are taken, i.e., B07: Bolton & Haehnelt (2007), FG08: Faucher-Giguère et al. (2008a), P09: Prochaska et al. (2009) and SC10: Songaila & Cowie (2010).

Redshift	$\lambda_{\text{mfp}}$ (cMpc)	Reference
$z = 2$	$909 \pm 252$	FG08, SC10
$z = 3$	$337 \pm 170$	FG08, SC10
$z = 4$	$170 \pm 15.5$	FG08, P09
$z = 5$	$83.4 \pm 21.6$	B07, SC10

star-forming galaxies at  $z \gtrsim 3$ . In this section, we quantify the contribution of the ionizing photons that are produced in stars to the build-up of the UVB radiation in our simulations. In addition, we calculate the implied average escape fraction that is required to generate the observed UVB.

### 3.4.3.1 Generating the UVB

Our simulation box is smaller than the typical mean free path of ionizing photons,  $\lambda_{\text{mfp}}$  (see Table 3.1). Therefore, the IGM gas can receive stellar ionizing radiation from a region that is larger than the simulation box. We note that using periodic boundaries to propagate the stellar ionizing photons is not a good solution to resolve this issue. In addition to increasing the computational expense, using periodic boundaries for RT would require us to account for the cosmological redshifting of ionizing photons. Moreover, because of the small size of our box, stellar photons traveling along paths that are nearly parallel to a side of the box may never intersect an optically thick absorber, if periodic boundaries are being used for RT. However, we can correct for the small size of the simulation box by requiring consistency between the simulated UV intensities that local sources produce in the IGM and existing observations/models of the UVB. In order to do that, we assume an isotropic and homogeneous universe that is in photoionization equilibrium, and  $\lambda_{\text{mfp}}/(1+z) \ll c/H(z)$ , so that we can ignore evolution and redshifting during the travel time of the photons. Then, the volume-weighted mean ionizing flux,  $F_\star$ , from stars is given by:

$$F_\star = \int_0^\infty u_\star e^{-\frac{r}{\lambda_{\text{mfp}}}} dr = \bar{u}_\star \lambda_{\text{mfp}}, \quad (3.8)$$

where  $\bar{u}_\star$  is the photon production rate per unit comoving volume. Moreover, one can express the mean ionizing flux in equation (3.8) in terms of the volume-weighted mean ionizing flux produced by source that are inside the simulation



box,  $F_{\star}^{\text{in}}$

$$F_{\star}^{\text{in}} = \int_0^{\alpha L_{\text{box}}} u_{\star} e^{-\frac{r}{\lambda_{\text{mfp}}}} dr, \quad (3.9)$$

where  $\alpha \lesssim 1$  is a geometrical factor. In the absence of any absorption and for uniform and isotropic distributions of gas and sources,  $\alpha$  is set by the average distance between two random points inside a cube. For a cube with unit length this would yield  $\alpha = 0.66$  (Robbins, 1978). Based on the average distance between the low-density gas (e.g., SPH particles with  $n_{\text{H}} \lesssim 10^{-5} \text{ cm}^{-3}$ ) and sources (i.e., star-forming particles) in our simulations, we find that the average value of  $\alpha$  is 0.66 which varies mildly with redshift from  $\alpha_0 = 0.54$  at  $z = 0$  to  $\alpha_5 = 0.79$  at  $z = 5$ .

As Table 3.1 shows, our simulation box is much smaller than the mean free path of ionizing photons. Therefore, we can use  $L_{\text{box}} \ll \lambda_{\text{mfp}}$  in equation (3.9) and get:

$$F_{\star}^{\text{in}} \approx \bar{u}_{\star} \alpha L_{\text{box}}, \quad (3.10)$$

Using equation (3.8) and (3.10), the total stellar UVB radiation flux can thus be written as:

$$F_{\star} \approx \frac{\lambda_{\text{mfp}}}{\alpha L_{\text{box}}} F_{\star}^{\text{in}}. \quad (3.11)$$

Therefore, the volume-weighted photoionization rate due to radiation produced by stars in the simulation,  $\Gamma_{\star}^{\text{in}}$ , which is close to the median photoionization rate in low-density gas, can be used to calculate the implied UVB photoionization rate after correcting for the small box size of the simulations:

$$\Gamma_{\text{UVB},\star} \approx \frac{\lambda_{\text{mfp}}}{\alpha L_{\text{box}}} \Gamma_{\star}^{\text{in}}. \quad (3.12)$$

As mentioned earlier, our simulations show that the photoionization rate from local stellar radiation approaches a density independent rate at low densities. We take this photoionization rate as  $\Gamma_{\star}^{\text{in}}$ . In addition, we use a compilation of available Lyman-limit mean free path measurements at different redshifts from the literature (i.e., from Bolton & Haehnelt, 2007; Faucher-Giguère et al., 2008a; Prochaska et al., 2009; Songaila & Cowie, 2010; see Table 3.1 and O’Meara et al., 2013 and Fumagalli et al., 2013 for new measurements). After converting the Lyman-limit mean free paths into the typical mean free path of ionizing photons with our assumed stellar spectrum<sup>2</sup>, we use equation (3.12) to derive the implied UVB photoionization rate.

Figure 3.7 shows the predicted contribution of stellar radiation to the UVB (filled circles). The error bars reflect the quoted error in the mean free path measurements. For comparison, the observational measurement of the UVB from the Ly $\alpha$  effective opacity by Bolton & Haehnelt (2007) and the modeled

<sup>2</sup>Because the effective hydrogen ionization cross section of stellar ionizing photons that we use in this work,  $\bar{\sigma}_{\star} = 2.9 \times 10^{-18} \text{ cm}^2$ , their typical mean free paths are  $\sim 2$  times longer than the mean free path of Lyman-limit photons which have hydrogen ionizing cross sections  $\sigma_0 = 6.8 \times 10^{-18} \text{ cm}^2$ .

Haardt & Madau (2001) UVB photoionization rates are also shown by orange diamonds and green dashed curve, respectively. Both the observational measurements and modeled UVB intensities are in good agreement with our simulation results for  $z > 2$ . However, their UVB intensity is slightly higher than ours at  $z = 2$ . The reason for this could be the absence of radiation from quasars in our simulations.

### 3.4.3.2 Average escape fractions

The average star formation rate density in our simulations is in good agreement with the observed cosmic star formation rate (Schaye et al., 2010). Therefore, the good agreement between the UVB intensities in our simulation and the ones inferred from the observations suggests that at  $z \gtrsim 3$  the average escape fractions of stellar ionizing photons in our simulations are also reasonable. However, as we will discuss in Appendix B2, the structure of the ISM is unresolved in our simulations. Therefore, the fact that the produced escape fractions are reasonable may be coincidental.

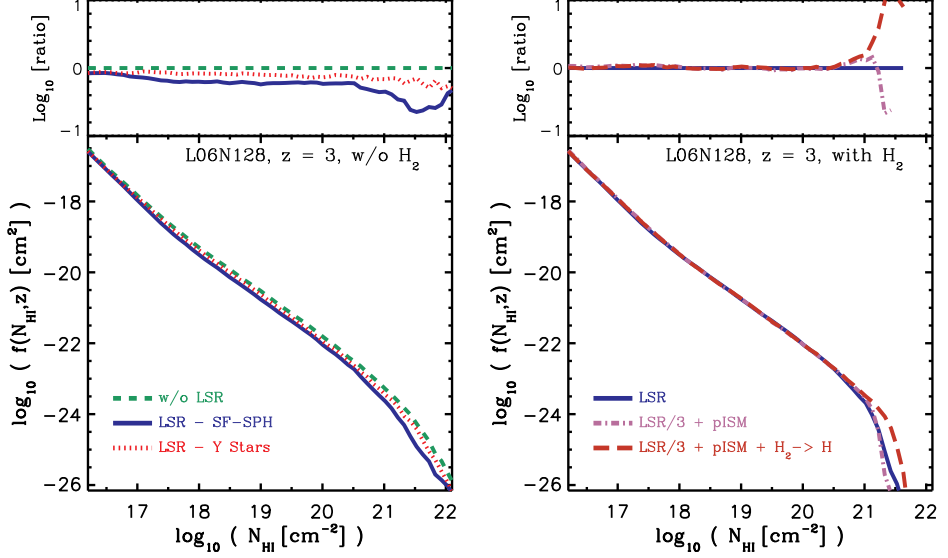
Since in our simulations both the intensity of stellar radiation in the IGM and the photon production rate are known, we can measure the mean star formation rate-weighted escape fraction of ionizing photons from galaxies into the IGM (see Appendix C):

$$\begin{aligned}
 f_{\text{esc}} &\sim 10^{-2} \left( \frac{\Gamma_{\star}^{\text{in}}}{10^{-14} \text{ s}^{-1}} \right) \left( \frac{\bar{\sigma}_{\star}}{2.9 \times 10^{-18} \text{ cm}^2} \right)^{-1} \\
 &\times \left( \frac{\dot{\rho}_{\star}}{0.15 \text{ M}_{\odot} \text{ yr}^{-1} \text{ cMpc}^{-3}} \right)^{-1} \\
 &\times \left( \frac{\alpha}{0.7} \right)^{-1} \left( \frac{L_{\text{box}}}{10 \text{ cMpc}} \right)^{-1} \left( \frac{1+z}{4} \right)^{-2}. \quad (3.13)
 \end{aligned}$$

The implied escape fractions are shown in the right panel of Figure 3.7. The simulated escape fractions are  $10^{-2} < f_{\text{esc}} < 10^{-1}$  between  $z = 2$  and 5 and they decrease with decreasing redshift below  $z = 4$ . This result is consistent with previous observational and theoretical studies (e.g., Shapley et al., 2006; Schaye, 2006; Gnedin et al., 2008; Kuhlen & Faucher-Giguère, 2012).

### 3.4.4 The impact of local stellar radiation on the H I column density distribution

The observed distribution of neutral hydrogen is often quantified by measuring the distribution of H I absorbers with different strengths in the spectra of background quasars (e.g., Kim et al., 2002; Péroux et al., 2005; O’Meara et al., 2007; Noterdaeme et al., 2009; Prochaska et al., 2009; Prochaska & Wolfe, 2009; O’Meara et al., 2013; Noterdaeme et al., 2012). The H I column density distribution function (CDDF) is defined as the number of systems at a given column



**Figure 3.8:** *Left:* The HI column density distribution function (CDDF) at  $z = 3$  with different ionizing sources. The blue solid curve shows our reference simulation which includes the UVB, local stellar radiation (LSR) and recombination radiation and the green dashed curve indicates the simulation without local stellar radiation (i.e., with the UVB and recombination radiation). While in the reference simulation star-forming SPH particles are used as ionizing sources, the HI CDDF that is shown with the red dotted curve indicates a simulation in which young stellar particles are used as sources. Using star-forming SPH particles as sources lowers the HI CDDF by  $\approx 0.5$  dex for  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ . However, using young stellar particles, which results in sampling issues (see §3.4.2), has a weak impact on the HI CDDF. For calculating the HI CDDF, the neutral gas is assumed to be fully atomic (i.e., no  $\text{H}_2$ ). *Right:* The HI CDDF at  $z = 3$  in the presence of local stellar radiation (i.e., star-forming SPH particles as sources) for different assumptions about the ISM. The blue solid curve shows our reference simulation in which all H atoms contribute to the absorption but molecular hydrogen does not contribute to the HI CDDF. The orchid dot-dashed curve shows a porous ISM model (see the text) where molecular hydrogen does not absorb ionizing radiation during the RT calculation (it is assumed to have a very small covering fraction). In order to reproduce the observed UVB intensity in this model, the local stellar radiation has been reduced by a factor of 3 compared to the fiducial model. The red long-dashed curve is identical to the orchid dot-dashed curve (porous ISM) but molecular hydrogen is assumed to dissociate into atomic hydrogen before calculating the HI CDDF. At higher HI column densities ( $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ ) the HI CDDF is highly sensitive to the assumptions about the unresolved ISM.

density, per unit column density, per unit absorption length,  $dX$ :

$$f(N_{\text{HI}}, z) \equiv \frac{d^2n}{dN_{\text{HI}}dX} \equiv \frac{d^2n}{dN_{\text{HI}}dz} \frac{H(z)}{H_0} \frac{1}{(1+z)^2}. \quad (3.14)$$

In order to study the effect of local stellar radiation on the H I CDDF, we project our simulation boxes on a two-dimensional grid and use this to calculate the column densities (see [Rahmati et al., 2013](#) for more details).

#### 3.4.4.1 Local stellar radiation and the H I CDDF at $z = 3$

The simulated H I CDDF at  $z = 3$  is shown in the left panel of [Figure 3.8](#) for different ionizing sources. The blue solid curve shows the H I CDDF in our reference simulation which includes local stellar radiation, the UVB and recombination radiation. For comparison, the H I CDDF without local stellar radiation (i.e., only including the UVB and recombination radiation) is shown with the green dashed curve. As the ratio between these two H I CDDFs in the top section of the left panel in [Figure 3.8](#) illustrates, the effect of local stellar radiation increases with H I column density and reaches a  $\sim 0.5$  dex reduction at  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ . For H I column densities lower than  $10^{17} \text{ cm}^{-2}$  on the other hand, the H I CDDF is insensitive to local stellar radiation. These trends are consistent with previous analytic arguments ([Miralda-Escudé, 2005](#); [Schaye, 2006](#)) and numerical simulations performed by [Fumagalli et al. \(2011\)](#).

As we discussed in [§3.4.2](#), using star-forming SPH particles as ionizing sources (i.e., our reference model) results in a better sampling than using young stellar particles as sources. We showed that if one uses young stellar particles as ionizing sources, the impact of local stellar radiation will be under-estimated in a large fraction of low-mass haloes in our simulations (see [Figure 3.6](#) and [3.5](#)). To illustrate this difference, the H I CDDF for the simulation in which young stellar particles are used, is shown with the red dotted curve in the left panel of [Figure 3.8](#). Indeed, the impact of local sources on the H I CDDF is much weaker if we use young stellar particles as ionizing sources. This may partly explain why [Yajima et al. \(2012\)](#) found that local sources did not affect the H I CDDF significantly.

#### 3.4.4.2 The impact of the unresolved ISM

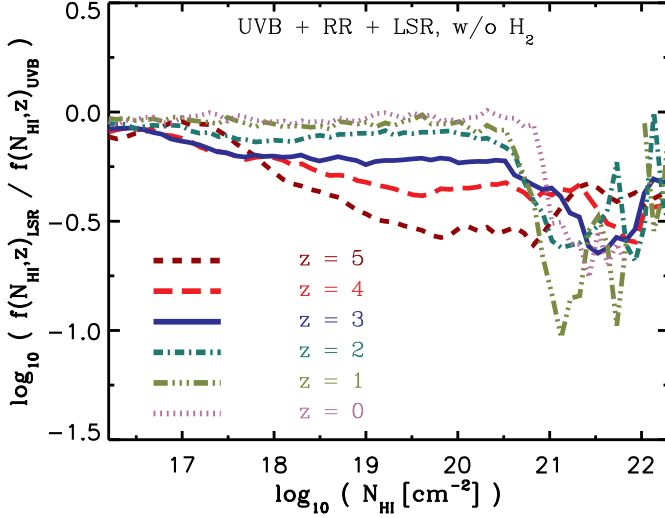
In our simulations, the ISM is modeled by enforcing a polytropic equation of state on SPH particles with densities  $n_{\text{H}} > 10^{-1} \text{ cm}^{-3}$ . This means that our simulations do not include a cold ( $T \ll 10^4 \text{ K}$ ) interstellar gas phase. This simple ISM modeling introduces uncertainties in the hydrogen neutral fraction calculations of the dense regions in our simulations. In addition, for gas with  $n_{\text{H}} \gg 10^{-2} \text{ cm}^{-3}$  the mean free path of ionizing photons is unresolved (see [Appendix B1](#)).

To estimate the impact of the structure of the ISM on the H I CDDF and the effect of local sources, we introduce a porous ISM model in which we assume that molecular hydrogen is confined to clouds with such small covering fractions that we can ignore them when performing the RT. Following Rahmati et al. (2013), we use the observationally inferred pressure law of Blitz & Rosolowsky (2006) to compute the molecular fraction (see Appendix A), but now we not only subtract the H<sub>2</sub> fraction when projecting the H I distribution to compute the H I CDDF, we also subtract it before doing the RT calculation (note that in the fiducial case we assumed H<sub>2</sub> fractions to be zero during the RT calculation). The conversion of diffuse atomic hydrogen into compact molecular clouds which do not absorb ionizing photons should facilitate the propagation of photons from star-forming regions into the IGM. Indeed, we find that replacing our reference uniform ISM model with the porous ISM model increases the resulting UVB photoionization rate by a factor of  $\sim 3$ . Therefore, we decreased the photon production rate of local stellar radiation by a factor of 3, such that the model with a porous ISM yields the same UVB intensity that is generated by our reference simulation and is in agreement with the observed UVB. This factor of 3 reduction could be interpreted as reflecting absorptions in molecular clouds that are hosting young stars<sup>3</sup>. One should note that despite the enforced agreement between the generated UVBs, the stellar photoionization rates at high and intermediate densities are different in the two simulations.

The right panel of Figure 3.8 shows that for  $N_{\text{HI}} < 10^{21} \text{ cm}^{-2}$  the H I CDDF in the simulation with a porous ISM (orchid dot-dashed curve) is almost identical to the H I CDDF in our reference simulation (blue solid curve). This suggests that for these column densities the predicted impact of local stellar radiation on the H I CDDF is robust to uncertainties regarding the small scale structure of the ISM. Instead, its impact is controlled by the amount of radiation that is propagating through the properly resolved intermediate densities towards the IGM, which is constrained by the observed UVB intensity. Note, however, that the impact of local sources on the high H I column density part of the distribution (i.e.,  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ ) is in fact highly dependent on the assumptions that are made about the physics of the ISM.

It is also possible that the intense stellar ionizing radiation within the ISM will effectively increase the H I column densities by dissociating hydrogen molecules. However, we implicitly neglected this effect in our simulation with a porous ISM by assuming that molecular clouds are not affected by local stellar radiation. To put an upper limit on the impact of this effect, one can assume that all molecular clouds are completely dissociated by the absorption of stellar radiation. The result of this exercise is shown by the red long-dashed curve in Figure 3.8. This shows that accounting for H<sub>2</sub> dissociation could reduce (or even reverse) the impact of local stellar radiation only at the very high column density end of the H I CDDF (i.e.,  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ ).

<sup>3</sup>Note that this is not equivalent to  $f_{\text{esc}} = 1/3$ . The radiation that leaves star-forming regions is still subject to significant absorption before it reaches the IGM.



**Figure 3.9:** The ratio between the H I CDDF with and without local stellar radiation at different redshifts. For calculating the H I CDDF, the neutral gas is assumed to be fully atomic (i.e., no H<sub>2</sub>). The impact of local stellar radiation decreases the H I CDDF by up to 1 dex. The impact of local stellar radiation on LLSs increases with redshift.

We stress that none of our ISM models are realistic. However, by considering very different models, we can nevertheless get an idea of the possible impact of our simplified treatment of the ISM and we conclude that our results are relatively robust for  $N_{\text{HI}} \ll 10^{21} \text{ cm}^{-2}$ .

### 3.4.4.3 Evolution

The evolution of the impact of local stellar radiation on the H I CDDF is illustrated in Figure 3.9 for our fiducial model. Each curve in this figure shows the ratio between the H I CDDF with and without local stellar radiation at a given redshift. To avoid the uncertainties about the conversion of atomic gas into H<sub>2</sub>, the H I CDDFs are computed assuming that the neutral gas is fully atomic (i.e., no H<sub>2</sub>). For all redshifts local stellar radiation has only a very small impact on the H I CDDF for  $N_{\text{HI}} \ll 10^{17} \text{ cm}^{-2}$  but significantly reduces the abundance of systems with  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ . The impact of local stellar radiation increases with redshift for  $10^{18} < N_{\text{HI}} < 10^{21} \text{ cm}^{-2}$ . While local sources significantly reduce the H I CDDF for  $N_{\text{HI}} \gg 10^{17} \text{ cm}^{-2}$  at  $z = 5$ , their effect only becomes significant for  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$  at  $z = 0$ . This might be attributed to decrease in the proper sizes of galaxies with redshift.

In Rahmati et al. (2013) we used simulations that include only the UVB and recombination radiation to show that for  $10^{18} \lesssim N_{\text{HI}} \lesssim 10^{20} \text{ cm}^{-2}$ , the H I CDDF

does not evolve at  $z \leq 3$  and increases with increasing redshift for  $z > 3$ . Figure 3.10 shows that if we also include local stellar radiation, the weak evolution of the H I CDDF in the Lyman Limit range extends to even higher redshifts (i.e.,  $z \lesssim 4$ ).

The predicted H I CDDFs are compared to observations in Figure 3.10. We note that the H I CDDF predicted by our reference simulation is not converged with respect to the size of the simulation box (see Appendix B in Rahmati et al., 2013). Since the photoionization caused by local stellar radiation only exceeds the UVB photoionization rate close to galaxies (see Figure 3.4), we can assume that the impact of local stellar radiation on the H I CDDF is independent of the size of the simulation box. Therefore, for all the curves shown in Figure 3.10 we have corrected the box size effect by multiplying the CDDF of the fiducial  $L = 6.25h^{-1}\text{Mpc}$  box by the ratio of the CDDF in a converged simulation (with  $L = 50h^{-1}\text{Mpc}$ ) and in our reference simulations with  $L = 6.25h^{-1}\text{Mpc}$ , both in the absence of local stellar radiation (i.e., with the UVB and recombination radiation).

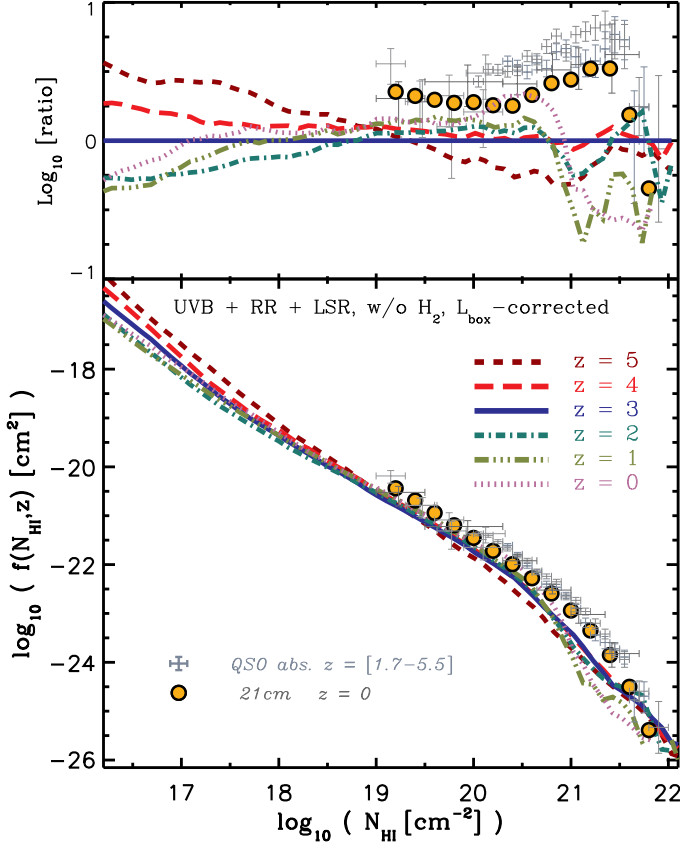
As we showed in Rahmati et al. (2013), the simulated H I CDDFs in the presence of the UVB and recombination radiation is in a reasonable agreement with observations (see also Altay et al., 2011) with a small deviation of  $\sim 0.2$  dex. But as the top section of Figure 3.10 illustrates, the addition of local stellar radiation increases this deviation to  $\gtrsim 0.5$  dex. While the difference between the predicted H I CDDFs and observations is larger at the highest H I column densities (i.e.,  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ ), we have seen that in this regime the effect of local sources is sensitive to the complex physics of the ISM, which our simulations do not capture. These very high H I column densities are also sensitive to the strength and the details of different feedback mechanisms (see Altay et al. in prep.). The small but significant discrepancy in the LL and weak DLA regime is therefore more interesting. It is important to confirm this discrepancy with larger simulations, so that a correction for box size will no longer be necessary, but this requires more computing power than is presently available to us.

### 3.5 Discussion and conclusions

The column density distribution function (CDDF) of neutral hydrogen inferred from observations of quasar absorption lines is the most accurately determined observable of the distribution of gas in the high-redshift Universe. Moreover, the high column density absorbers ( $N_{\text{HI}} > 10^{17} \text{ cm}^{-2}$ ) arise in gas that is either already part of the ISM or will soon accrete onto a galaxy (van de Voort et al., 2012) and hence these systems directly probe the fuel for star formation.

To predict the distribution of high column density absorbers, it is necessary to combine cosmological hydrodynamical simulations with accurate radiative transfer (RT) of ionizing radiation. Because of the cost and complexity associated with RT calculations that include many sources, nearly all studies have only





**Figure 3.10:** The H<sub>I</sub> CDDF at different redshifts. Curves show predictions for the reference simulation in the presence of the UVB, diffuse recombination radiation and local stellar radiation after correcting for the box size (see the text). For calculating the H<sub>I</sub> CDDF, the neutral gas is assumed to be fully atomic (i.e., no H<sub>2</sub>). The observational data points represent a compilation of various quasar absorption line observations at high redshifts (i.e.,  $z = [1.7, 5.5]$ ) taken from Péroux et al. (2005) with  $z = [1.8, 3.5]$ , O’Meara et al. (2007) with  $z = [1.7, 4.5]$ , Noterdaeme et al. (2009) with  $z = [2.2, 5.5]$  and Prochaska & Wolfe (2009) with  $z = [2.2, 5.5]$ . The orange filled circles show the best-fit based on the low-redshift 21 cm observations of Zwaan et al. (2005). The *top section* shows the ratio between the H<sub>I</sub> CDDFs at different redshifts and the predicted CDDF at  $z = 3$ . While simulations without local stellar radiation are in reasonable agreement with the observed H<sub>I</sub> CDDF (not shown here, but see Rahmati et al., 2013), in the presence of local stellar radiation the simulated H<sub>I</sub> CDDFs of LLSs deviate from observations by a factor of  $\approx 2$ . Local stellar radiation weakens the evolution of the H<sub>I</sub> CDDFs at  $10^{17} \lesssim N_{\text{HI}} \lesssim 10^{20} \text{ cm}^{-2}$ .



considered ionization by the ultraviolet background radiation (UVB), whose intensity can be inferred from observations of the Ly $\alpha$  forest. However, analytic arguments suggest that the radiation field to which high column density absorbers are exposed is typically dominated by local stellar sources (Miralda-Escudé, 2005; Schaye, 2006). It is therefore important to investigate whether the remarkable success of simulations that consider only the UVB, such as the agreement with the observed CDDF over 10 orders of magnitude in column density at  $z = 3$  (Altay et al., 2011) as well as its evolution down to  $z = 0$  (Rahmati et al., 2013), is compromised when local sources are included. Here we addressed this question by repeating some of the simulations of Rahmati et al. (2013) with our RT code TRAPHIC (Pawlik & Schaye, 2008, 2011; Rahmati et al., 2013), but this time including not only the UVB and diffuse recombination radiation, but also local stellar sources.

In agreement with the analytic predictions of Miralda-Escudé (2005) and Schaye (2006), we found that local stellar radiation is unimportant for  $N_{\text{HI}} \ll 10^{17} \text{ cm}^{-2}$  and dominates over the UVB for high column density absorbers. For all redshifts considered here (i.e.,  $z \leq 5$ ), local sources strongly reduce the abundance of systems with  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ . The impact of local sources increases with redshift for  $10^{18} < N_{\text{HI}} < 10^{21} \text{ cm}^{-2}$ . At  $z = 5$  the CDDF is substantially reduced for  $N_{\text{HI}} \gg 10^{17} \text{ cm}^{-2}$ , but at  $z = 0$  the effect only becomes significant for  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ . As a result, the remarkable lack of evolution in the CDDF that we found in Rahmati et al. (2013) for  $z = 0 - 3$ , and which is also observed, extends to  $z = 4$  if local sources are taken into account. On the other hand, the agreement with the observed  $z \sim 3$  CDDF is not quite as good as before, with the simulations underpredicting the rates of incidence of  $10^{19} < N_{\text{HI}} < 10^{21} \text{ cm}^{-2}$  absorbers by factors of a few. However, because of the large corrections that we had to make because of the small size of the simulation box used to study the effect of local sources ( $6.25h^{-1}\text{Mpc}$ ), this discrepancy will have to be confirmed with larger simulations. Moreover, we did not account for possible hydrodynamical effects that might be caused by extra heating due to ionizing radiation from local sources. This process might change the distribution of gas that is affected by local stellar radiation and requires further investigation.

We found that the average photoionization rate due to young stars in high-density gas is weakly dependent on the gas density and is  $\sim 10^{-13} \text{ s}^{-1}$ . We showed analytically that this rate follows directly from the imposed (and observed) Kennicutt-Schmidt star formation law if we assume that most of the ionizing photons that are produced by star-forming gas are absorbed on scales  $\lesssim \text{kpc}$ . However, in reality we expect the photoionization rate in the ISM to fluctuate more strongly than predicted by simulations like ours, which lack the resolution required to model the cold, interstellar phase.

Indeed, the spatial resolution that is required for accurate RT of ionizing radiation through the ISM is several orders of magnitude higher than the smallest scales accessible in current cosmological simulations. This makes tackling this problem in the near future hardly feasible and poses a difficult challenge for

studying the impact of local stellar radiation on the distribution of H I in and around galaxies. Fortunately, one can circumvent part of the problem by tuning the production rate of ionizing photons (which is equivalent to adjusting the escape fraction of ionizing photons from the unresolved ISM) such that the models reproduce the observed mean photoionization rate in the IGM (after subtracting the contribution from quasars). In other words, if one knows the amount of ionizing radiation that is required to reach the IGM, its ionization impact on the intervening gas can be determined even if we cannot predict what fraction escapes from the immediate vicinity of the young stars.

We adopted this approach but found that tuning was unnecessary for our reference simulation at  $z \sim 3$ . Moreover, since our simulations also yield star formation histories that are in good agreement with observations (Schaye et al., 2010), we used them to constrain the implied star formation rate weighted mean escape fraction that relates the predicted star formation rate density to the intensity of the UVB. We found that the average escape fraction in our simulations is  $10^{-2} - 10^{-1}$  at  $z = 2 - 5$ , which agrees with previous constraints on the escape fraction from observations (e.g., Shapley et al., 2006) and theoretical work (e.g., Schaye, 2006; Gnedin et al., 2008; Kuhlen & Faucher-Giguère, 2012).

The limited spatial resolution of cosmological simulations mandates the use of simplified models for the structure of the ISM. To estimate the impact of such subgrid models on the CDDF and on the effect of local sources, we varied some of the underlying assumptions. In particular, we considered a porous ISM model which assumes that molecular hydrogen is confined to clouds with such small covering fractions that we can ignore them when performing the RT. We also considered a model which assumes that all molecular clouds are completely dissociated, but not ionised, by the absorption of stellar radiation. Although none of these models are realistic, we used them to estimate the potential impact of our simplified treatment of the ISM. We found that provided that we rescale the source luminosities so that the different models all reproduce the observed background radiation, the models predict nearly the same H I CDDFs in the regime where the absorbers are well resolved. We therefore concluded that our results on the effect of local sources are relatively robust for  $N_{\text{HI}} \ll 10^{21} \text{ cm}^{-2}$ , but that their predicted impact is highly sensitive to the assumptions about the ISM for  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ .

Different studies have found qualitatively different results for the impact of local stellar radiation on the CDDF. For instance, Fumagalli et al. (2011) used relatively high-resolution zoomed simulations of individual objects to demonstrate that local stellar radiation significantly reduces the abundance of high column density absorbers. On the other hand, some studies using cosmological simulations similar to those presented here (with roughly the same resolution) found that local stellar radiation has a negligible impact on the CDDF (Nagamine et al., 2010; Yajima et al., 2012).

We found that difference in the resolutions of the simulations that were used in these previous studies may explain their inconsistent findings. Using star

particles as sources, as was done by (Nagamine et al., 2010; Yajima et al., 2012), we also found that local sources have a negligible impact on the abundance of strong HI systems. However, we demonstrated that it is possible to dramatically improve the sampling of the distribution of ionizing sources by using star-forming gas particles (i.e., gas with densities at least as high as those typical of the warm ISM), thus effectively increasing the resolution without modifying the time-averaged production rate of ionizing photons. We adopted this strategy in our fiducial models and found, as summarized above, that the radiation from local sources significantly affects the high column density end of the CDDF. This result is in agreement with Fumagalli et al. (2011), Gnedin (2010) and analytic estimates of Miralda-Escudé (2005) and Schaye (2006), and confirms that poor sampling of the distribution of ionizing sources can lead to an under-estimation of the impact of local stellar radiation.

Further progress will require higher resolution simulations and, most importantly, more realistic models for the ISM. In the near future it will remain unfeasible to accomplish this in a cosmologically representative volume. Until this challenge is met, predictions for the escape fractions of ionizing radiation averaged over galaxy populations should be considered highly approximate. Predictions for the abundances of LL and weak DLA systems based on models that neglect local sources of stellar radiation should be interpreted with care, particularly for  $z > 2$ . Predictions for the CDDF in the strong DLA regime ( $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ ) must be considered highly approximate at all redshifts.

## Acknowledgments

We thank the anonymous referee for a helpful report. We also would like to thank Dušan Kereš, J. Xavier Prochaska and Tom Theuns for valuable discussions. The simulations presented here were run on the Cosmology Machine at the Institute for Computational Cosmology in Durham (which is part of the DiRAC Facility jointly funded by STFC, the Large Facilities Capital Fund of BIS, and Durham University) as part of the Virgo Consortium research programme. This work was sponsored with financial support from the Netherlands Organization for Scientific Research (NWO), also through a VIDI grant and an NWO open competition grant. We also benefited from funding from NOVA, from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement 278594-GasAroundGalaxies and from the Marie Curie Training Network CosmoComp (PITN-GA-2009-238356). AHP receives funding from the European Union’s Seventh Framework Programme (FP7/2007-2013) under grant agreement number 301096-proFeSsOR.

## References

- Aguirre, A., Dow-Hygelund, C., Schaye, J., & Theuns, T. 2008, *ApJ*, 689, 851
- Altay, G., Theuns, T., Schaye, J., Crighton, N. H. M., & Dalla Vecchia, C. 2011, *ApJL*, 737, L37
- Bigiel, F., Leroy, A., Walter, F., et al. 2008, *AJ*, 136, 2846
- Bird, S., Vogelsberger, M., Sijacki, D., et al. 2013, *MNRAS*, 514
- Blitz, L., & Rosolowsky, E. 2006, *ApJ*, 650, 933
- Bolton, J. S., Haehnelt, M. G., Viel, M., & Springel, V. 2005, *MNRAS*, 357, 1178
- Bolton, J. S., & Haehnelt, M. G. 2007, *MNRAS*, 382, 325
- Cen, R., Ostriker, J. P., Prochaska, J. X., & Wolfe, A. M. 2003, *ApJ*, 598, 741
- Chabrier, G. 2003, *PASP*, 115, 763
- Croft, R. A. C. 2004, *ApJ*, 610, 642
- Dalla Vecchia, C., & Schaye, J. 2008, *MNRAS*, 387, 1431
- Faucher-Giguère, C.-A., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008, *ApJ*, 688, 85
- Fumagalli, M., Prochaska, J. X., Kasen, D., et al. 2011, *MNRAS*, 418, 1796
- Fumagalli, M., O'Meara, J. M., Prochaska, J. X., & Worseck, G. 2013, *arXiv:1308.1101*
- Gardner, J. P., Katz, N., Hernquist, L., & Weinberg, D. H. 1997, *ApJ*, 484, 31
- Gnedin, N. Y., Kravtsov, A. V., & Chen, H.-W. 2008, *ApJ*, 672, 765
- Gnedin, N. Y. 2010, *ApJL*, 721, L79
- Haardt F., Madau P., 2001, in *Clusters of Galaxies and the High Redshift Universe Observed in X-rays*, Neumann D. M., Tran J. T. V., eds.
- Haehnelt, M. G., Steinmetz, M., & Rauch, M. 1998, *ApJ*, 495, 647
- Haehnelt, M. G., Madau, P., Kudritzki, R., & Haardt, F. 2001, *ApJL*, 549, L151
- Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007, *ApJ*, 654, 731
- Katz, N., Weinberg, D. H., Hernquist, L., & Miralda-Escude, J. 1996, *ApJL*, 457, L57
- Kennicutt, R. C., Jr. 1998, *ApJ*, 498, 541
- Kim, J.-h., Krumholz, M. R., Wise, J. H., et al. 2012, *arXiv:1210.3361*
- Kim, T.-S., Carswell, R. F., Cristiani, S., D'Odorico, S., & Giallongo, E. 2002, *MNRAS*, 335, 555
- Komatsu, E., et al. 2011, *ApJS*, 192, 18
- Kroupa, P. 2001, *MNRAS*, 322, 231
- Kuhlen, M., & Faucher-Giguère, C.-A. 2012, *MNRAS*, 423, 862
- Leitherer, C., Schaerer, D., Goldader, J. D., et al. 1999, *ApJS*, 123, 3
- McQuinn, M., Oh, S. P., & Faucher-Giguère, C.-A. 2011, *ApJ*, 743, 82
- Miralda-Escudé, J. 2005, *ApJL*, 620, L91
- Nagamine, K., Springel, V., & Hernquist, L. 2004, *MNRAS*, 348, 421
- Nagamine, K., Choi, J.-H., & Yajima, H. 2010, *ApJL*, 725, L219
- Noterdaeme, P., Petitjean, P., Ledoux, C., & Srianand, R. 2009, *A&A*, 505, 1087
- Noterdaeme, P., Petitjean, P., Carithers, W. C., et al. 2012, *arXiv:1210.1213*
- O'Meara, J. M., Prochaska, J. X., Burles, S., et al. 2007, *ApJ*, 656, 666

- O'Meara, J. M., Prochaska, J. X., Worseck, G., Chen, H.-W., & Madau, P. 2013, *ApJ*, 765, 137
- Osterbrock, D. E., & Ferland, G. J. 2006, *Astrophysics of gaseous nebulae and active galactic nuclei*, 2nd. ed. by D.E. Osterbrock and G.J. Ferland. Sausalito, CA: University Science Books, 2006,
- Paardekooper, J.-P., Pelupessy, F. I., Altay, G., & Kruip, C. J. H. 2011, *A&A*, 530, A87
- Pawlik, A. H., & Schaye, J. 2008, *MNRAS*, 389, 651
- Pawlik, A. H., & Schaye, J. 2011, *MNRAS*, 412, 1943
- Péroux, C., Dessauges-Zavadsky, M., D'Odorico, S., Sun Kim, T., & McMahon, R. G. 2005, *MNRAS*, 363, 479
- Pontzen, A., Governato, F., Pettini, M., et al. 2008, *MNRAS*, 390, 1349
- Prochaska, J. X., & Wolfe, A. M. 2009, *ApJ*, 696, 1543
- Prochaska, J. X., Worseck, G., & O'Meara, J. M. 2009, *ApJL*, 705, L113
- Rahmati, A., Pawlik, A. H., Raičević, M., & Schaye, J. 2013a, *MNRAS*, 430, 2427
- Razoumov, A. O., Norman, M. L., Prochaska, J. X., & Wolfe, A. M. 2006, *ApJ*, 645, 55
- Robbins, D. 1978, *Amer. Math. Monthly* 85, 278
- Salpeter, E. E. 1955, *ApJ*, 121, 161
- Schaye, J. 2001, *ApJ*, 559, 507
- Schaye, J. 2004, *ApJ*, 609, 667
- Schaye, J. 2006, *ApJ*, 643, 59
- Schaye, J., & Dalla Vecchia, C. 2008, *MNRAS*, 383, 1210
- Schaye, J., Dalla Vecchia, C., Booth, C. M., et al. 2010, *MNRAS*, 402, 1536
- Shapley, A. E., Steidel, C. C., Pettini, M., Adelberger, K. L., & Erb, D. K. 2006, *ApJ*, 651, 688
- Songaila, A., & Cowie, L. L. 2010, *ApJ*, 721, 1448
- Springel, V. 2005, *MNRAS*, 364, 1105
- Vanzella, E., Giavalisco, M., Inoue, A. K., et al. 2010, *ApJ*, 725, 1011
- van de Voort, F., Schaye, J., Altay, G., & Theuns, T. 2012, *MNRAS*, 421, 2809
- Wiersma, R. P. C., Schaye, J., Theuns, T., Dalla Vecchia, C., & Tornatore, L. 2009a, *MNRAS*, 399, 574
- Wiersma, R. P. C., Schaye, J., & Smith, B. D. 2009b, *MNRAS*, 393, 99
- Yajima, H., Choi, J.-H., & Nagamine, K. 2012, *MNRAS*, 427, 2889
- Zuo, L. 1992, *MNRAS*, 258, 36
- Zwaan, M. A., van der Hulst, J. M., Briggs, F. H., Verheijen, M. A. W., & Ryan-Weber, E. V. 2005, *MNRAS*, 364, 1467

## Appendix A: Hydrogen molecular fraction

To account for the effect of molecular hydrogen, we adopt the same  $H_2$  conversion relation that [Altay et al. \(2011\)](#) used to successfully reproduce the  $H\text{I}$  CDDF

high-end cut-off. We follow Blitz & Rosolowsky (2006) and adopt an observationally inferred scaling relation between the gas pressure and the ratio between molecular and total hydrogen surface densities:

$$R_{\text{mol}} = \left( \frac{P_{\text{ext}}}{P_0} \right)^\alpha, \quad (3.15)$$

where  $R_{\text{mol}} \equiv \Sigma_{\text{H}_2}/\Sigma_{\text{H I}}$ ,  $P_{\text{ext}}$  is the galactic mid-plane pressure,  $\alpha = 0.92$  and  $P_0/k_b = 3.5 \times 10^4 \text{ cm}^{-3} \text{ K}$ . Furthermore, if we assume  $R_{\text{mol}}$  gives also the local mass ratio between molecular and atomic hydrogen, the fraction of gas mass which is in molecular form can be written as

$$f_{\text{H}_2} = \frac{M_{\text{H}_2}}{M_{\text{TOT}}} = \frac{M_{\text{H}_2}}{M_{\text{H}_2} + M_{\text{H I}}} = \frac{1}{1 + R_{\text{mol}}^{-1}}. \quad (3.16)$$

The last equation also assumes that at high densities ionization is not dominant and the gas is either molecular or neutral. In our simulations we model the multiphase ISM by imposing an effective equation of state with pressure  $P \propto \rho^{\gamma_{\text{eff}}}$  for densities  $n_{\text{H}} > n_{\text{H}}^*$ , where  $n_{\text{H}}^* = 0.1 \text{ cm}^{-3}$  which is normalized to  $P_*/k_b = 1.08 \times 10^3 \text{ cm}^{-3} \text{ K}$  at the threshold. Therefore we have:

$$P_{\text{ext}} = P_* \left( \frac{n_{\text{H}}}{n_{\text{H}}^*} \right)^{\gamma_{\text{eff}}}. \quad (3.17)$$

To make the Jeans mass and the ratio of the Jeans length to the SPH kernel independent of the density, we use  $\gamma_{\text{eff}} = 4/3$  (Schaye & Dalla Vecchia, 2008). Consequently, after combining equations (3.15), (3.16) & (3.17) the fraction of gas mass which is converted into H<sub>2</sub>,  $f_{\text{H}_2}$  can be written as:

$$f_{\text{H}_2} = \left( 1 + A \left( \frac{n_{\text{H}}}{n_{\text{H}}^*} \right)^{-\beta} \right)^{-1}, \quad (3.18)$$

where  $A = (P_*/P_0)^{-\alpha} = 24.54$  and  $\beta = \alpha \gamma_{\text{eff}} = 1.23$ .

## Appendix B: Resolution effects

### B1: Limited spatial resolution at high densities

At high densities, ionizing photons are typically absorbed within a short distance from the sources. The propagation of ionizing photons in these regions is therefore controlled by distribution of gas on scales that are comparable to the very short mean free path of ionizing photons. In Figure 3.11, the mean free path of stellar ionizing photons,  $\lambda_{\text{mfp}} = 1/(n_{\text{H I}} \bar{\sigma}_*)^4$ , is plotted as a function of density for our reference simulation at  $z = 3$ . The blue dashed curve in Figure 3.11

---

<sup>4</sup> $\bar{\sigma}_*$  is the spectrally averaged hydrogen photoionization cross section for stellar ionizing photons. For the spectral shape of the stellar radiation, we adopt a blackbody spectrum with  $T_{\text{bb}} = 5 \times 10^4 \text{ K}$ .

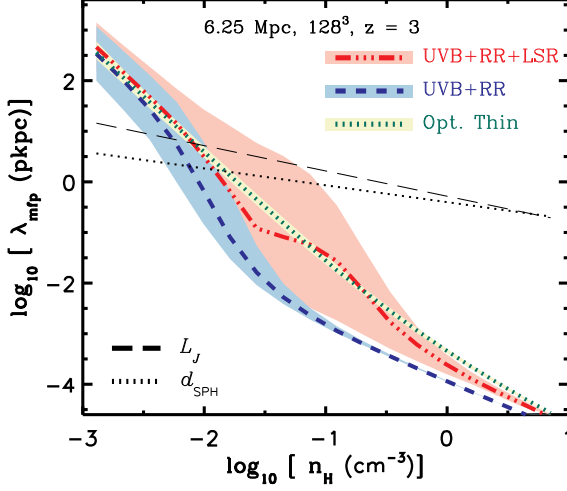


shows the result when only the UVB and recombination radiation are included. The effect of adding local stellar radiation is shown by the red dot-dashed curve. For comparison, also the result for a constant UVB radiation (i.e., optically thin gas) is illustrated with the green dotted curve. The ionization by local stellar radiation decreases the hydrogen neutral fraction at  $n_{\text{H}} \gtrsim 10^{-2} \text{ cm}^{-3}$  which in turn increases the mean free path of ionizing photons making it comparable to the optically thin case.

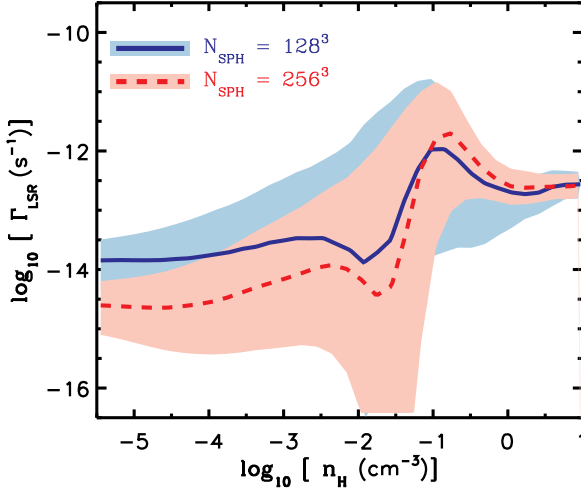
The simulations cannot provide any reliable information about the spatial distribution of gas and ionizing sources on scales smaller than the resolution limit (i.e., the typical distance between SPH particles) which is shown by the black dotted line in Figure 3.11. At densities for which the mean free path of ionizing photons is shorter than the spatial resolution, the RT results may not be accurate since all the photons that are emitted by sources are absorbed by their immediate neighbors. This happens at densities  $n_{\text{H}} \gtrsim 10^{-2} \text{ cm}^{-3}$  without local stellar radiation (blue dashed curve in Figure 3.11) and at densities  $n_{\text{H}} \gtrsim 10^{-1} \text{ cm}^{-3}$  with local stellar radiation (red dot-dashed curve in Figure 3.11). On the other hand, RT is irrelevant if the gas is highly neutral, which is predicted to be the case at slightly higher densities (see Figure 3.3).

## B2: The impact of a higher resolution on the RT

In Rahmati et al. (2013) we showed that the self-shielding limit is not very sensitive to the resolution of the simulation. On the other hand, the small scale structure of the ISM may significantly change the propagation of stellar radiation and their impact on the H I distribution. While this suggests that the impact of local stellar radiation might be sensitive to the resolution of the underlying simulation, our approach of tuning the escaped stellar radiation such that it can generate the desired UVB, circumvent most resolution effects. To study the impact of resolution on the propagation of stellar ionizing photons in the ISM and their escape to the IGM, we performed a simulation similar to our reference simulation but using 8 times more dark matter and SPH particles (i.e., using  $256^3$  dark matter particles and the same number of SPH particles whose masses are 8 times lower than in our reference simulation). Figure 3.12 shows that at  $z = 3$ , the photoionization rate of stellar radiation in the higher resolution simulation (red dashed curve) is qualitatively similar to that obtained from our reference simulation (blue solid curve). As expected, the stellar photoionization rate at the highest densities is in agreement with our analytic estimate in §3.2. However, because of the shorter inter-particle distances in the higher resolution simulation, the stellar photoionization rate peaks at slightly higher densities. The biggest difference between the stellar photoionization rates of the high-resolution and our reference simulation occurs at the lowest densities. Despite having a  $\sim 2$  times higher total star formation rate, the high-resolution simulation results in a UVB which has a  $\sim 5$  times lower photoionization rate. This means that the effective escape fraction in the high-resolution simulation is



**Figure 3.11:** The mean free path of stellar ionizing photons as a function of hydrogen number density for the reference simulation at  $z = 3$ . The red dot-dashed curve shows the result when all radiation sources, i.e., UVB, local stellar radiation (LSR) and recombination radiation (RR), are included and the blue dashed curve shows the result when local sources of radiation are not present. The green dotted curve shows the mean free path in a simulation that assumes a completely optically thin gas. The typical distance between SPH particles as a function of density is shown by the black dotted line. The Jeans length for a given density is illustrated by the black dashed line. The colored lines indicate the medians and shaded regions around them show the 15% – 85% percentiles.



**Figure 3.12:** Comparison between the stellar photoionization rate profiles in the reference simulation and a simulation with 8 times higher resolution at  $z = 3$ . The colored lines indicate the medians and shaded regions around them shows the 15% – 85% distributions.



$\sim 10$  times lower than for the reference simulation which has  $f_{\text{esc}} \sim 10^{-2}$ .

Finally, we emphasize that due to our simplified treatment of the ISM, it is not even clear whether the results become more realistic with increasing resolution.

## Appendix C: Calculation of the escape fraction

Using the equilibrium photon production rate per unit star formation rate from equation (3.2), the total production rate of ionizing photons in the simulation box,  $\dot{N}_{\gamma,*}$ , can be calculated as a function of the comoving star formation rate density within the simulation box,  $\dot{\rho}_*$  :

$$\dot{N}_{\gamma,*} = 2 \times 10^{53} \left( \frac{\dot{\rho}_* L_{\text{box}}^3}{1 M_{\odot} \text{ yr}^{-1}} \right). \quad (3.19)$$

Noting that the typical time the ionizing photons spend inside the simulation box is  $\sim \alpha L_{\text{box}} / (1+z)/c$ , and letting  $f_{\text{esc}}$  be the star formation rate-weighted mean fraction of ionizing photons that escape into the IGM, one can calculate the comoving equilibrium photon number density in the IGM:

$$\bar{n}_{\gamma} \sim \frac{f_{\text{esc}} \alpha \dot{N}_{\gamma,*}}{L_{\text{box}}^2 c (1+z)}, \quad (3.20)$$

where we ignored absorptions outside the host galaxies because  $L_{\text{box}} \ll \lambda_{\text{mfp}}$ .

The comoving number density of ionizing photons in the simulation box is related to the effective hydrogen photoionization rate they produce in the IGM,  $\Gamma_{*}^{\text{in}}$ :

$$\bar{n}_{\gamma} = \frac{\Gamma_{*}^{\text{in}}}{c \bar{\sigma}_{*} (1+z)^3}, \quad (3.21)$$

where  $\bar{\sigma}_{*}$  is the effective hydrogen ionization cross section for stellar photons. Combining equations (3.20) and (3.21) yields the effective escape fraction of ionizing photons from stars into the IGM:

$$f_{\text{esc}} \sim \frac{\Gamma_{*}^{\text{in}} L_{\text{box}}^2}{\bar{\sigma}_{*} \alpha \dot{N}_{\gamma,*} (1+z)^2}. \quad (3.22)$$

After putting numbers in equation (3.22) and using equation (3.19), the escape fraction becomes:

$$\begin{aligned} f_{\text{esc}} &\sim 10^{-2} \left( \frac{\Gamma_{*}^{\text{in}}}{10^{-14} \text{ s}^{-1}} \right) \left( \frac{\bar{\sigma}_{*}}{2.9 \times 10^{-18} \text{ cm}^2} \right)^{-1} \\ &\times \left( \frac{\dot{\rho}_*}{0.15 M_{\odot} \text{ yr}^{-1} \text{ cMpc}^{-3}} \right)^{-1} \\ &\times \left( \frac{\alpha}{0.7} \right)^{-1} \left( \frac{L_{\text{box}}}{10 \text{ cMpc}} \right)^{-1} \left( \frac{1+z}{4} \right)^{-2}. \end{aligned} \quad (3.23)$$



# 4

## PREDICTIONS FOR THE RELATION BETWEEN STRONG HI ABSORBERS AND GALAXIES AT REDSHIFT 3

We combine cosmological, hydrodynamical simulations with accurate radiative transfer corrections to investigate the relation between strong HI absorbers ( $N_{\text{HI}} > 10^{16} \text{ cm}^{-2}$ ) and galaxies at redshift  $z = 3$ . We find a strong anti-correlation between the column density and the impact parameter that connects the absorber to the nearest galaxy. The median impact parameters for Lyman Limit (LL) and Damped Lyman $\alpha$  (DLA) systems are  $\sim 10$  and  $\sim 1$  proper kpc, respectively. If normalized to the size of the halo of the nearest central galaxy, the median impact parameters for LL and DLA systems become  $\sim 1$  and  $\sim 0.1$  virial radii, respectively. At a given HI column density the impact parameter increases with the mass of the closest galaxy, in agreement with observations. We predict most strong HI absorbers to be most closely associated with extremely low-mass galaxies,  $M_{\star} \lesssim 10^8 M_{\odot}$ . We also find a correlation between the column density of absorbers and the mass of the nearest galaxy. This correlation is most pronounced for DLAs with  $N_{\text{HI}} > 10^{21} \text{ cm}^{-2}$  which are typically close to galaxies with  $M_{\star} \gtrsim 10^9 M_{\odot}$ . Similar correlations exist between column density and other properties of the associated galaxies such as their star formation rates, halo masses and HI content. The galaxies nearest to HI absorbers are typically far too faint to be detectable with current instrumentation, which is consistent with the high rate of (often unpublished) non-detections in observational searches for the galaxy counterparts of strong HI absorbers. Moreover, we predict that the detected nearby galaxies are typically not the galaxies that are most closely associated with the absorbers, thus causing the impact parameters, star formation rates and stellar masses of the counterparts to be biased high.

Alireza Rahmati, Joop Schaye  
*to be submitted*

## 4.1 Introduction

Studies of high-redshift galaxy are often based on the light emitted by stars and hot/ionized gas. This limits the observations to the small fraction of galaxies that are bright enough to be detected in emission. Given the large number density of faint galaxies, one may speculate that most high-redshift galaxies are missing from observational studies. The analysis of absorption features in the spectra of background QSOs, provides an alternative probe of the distribution of matter at high redshifts. The large distances that separate most absorbers from their background QSOs make it unlikely that there is a physical connection between them. This opens up a window to study an unbiased sample of matter that resides between us and background QSOs. The rare strong H I Ly $\alpha$  absorbers which are easily recognizable in the spectra of background QSOs due to their Ly $\alpha$  damping wings, for which they are called Damped Lyman- $\alpha$  (DLA) systems<sup>1</sup>, are of particular interest. DLAs are likely to be representative of the cold gas in, or close to, the interstellar medium (ISM) in high-redshift galaxies (Wolfe et al., 1986). Because of this, DLAs provide a unique opportunity to define an absorption-selected galaxy sample and to study the ISM, particularly at early stages of galaxy formation, and they have therefore been studied intensely since their discovery (see Wolfe et al., 2005 for a review).

Based on the observed velocity width of metal lines associated with DLAs, it was initially suggested that large, massive galactic disks are responsible for the observed DLAs at  $z \sim 3$  (Prochaska & Wolfe, 1997, 1998). However, it has been shown that (collections of) smaller systems are also capable of having high velocity dispersions as a result of infall of material during structure formation (Haehnelt et al., 1998) or galactic winds (McDonald & Miralda-Escudé, 1999; Schaye, 2001a). Nevertheless, reproducing the observed velocity width distribution remains a challenge for hydrodynamical simulations (e.g., Razoumov et al., 2006; Pontzen et al., 2008).

Some recent studies suggest that at  $z \sim 2 - 3$ , a large fraction of strong H I absorbers like Lyman Limit Systems (LLS;  $N_{\text{HI}} > 10^{17} \text{ cm}^{-2}$ ) and DLAs are primarily associated with galaxies similar to Lyman-Break Galaxies (e.g., Steidel et al., 2010; Rudie et al., 2012; Font-Ribera et al., 2012), which have stellar and total halo masses  $\sim 10^{10}$  and  $\sim 10^{12} M_{\odot}$ , respectively. If such massive galaxies were indeed the prime hosts of strong H I absorbers, then many of the galaxy counterparts of strong absorbers should be detectable with current surveys. However, observations that aim to find galaxies close to DLAs often result in non-detections (e.g., Foltz et al., 1986; Smith et al., 1989; Lowenthal et al., 1995; Bunker et al., 1999; Prochaska et al., 2002; Kulkarni et al., 2006; Rahmani et al., 2010; Bouché et al., 2012) or find galaxies that are at unexpectedly large impact parameters from DLAs (e.g., Yanny et al., 1990; Teplitz et al., 1998; Mannucci et al., 1998). Those findings suggest that strong H I

---

<sup>1</sup>The official column density limit of a DLA is somewhat arbitrarily defined to be  $N_{\text{HI}} > 10^{20.3} \text{ cm}^{-2}$ .

systems such as DLAs are more closely associated to low-mass galaxies which are too faint to be observable with the detection thresholds of the current studies.

Because observational studies are limited by the small number of known DLAs and are missing low-mass galaxies, we have to resort to cosmological simulations to help us understand the link between DLAs and galaxies. Many studies have used simulations to investigate the nature of strong H I absorbers and particularly DLAs (e.g., Gardner et al., 1997, 2001; Haehnelt et al., 1998; Nagamine et al., 2004; Razoumov et al., 2006; Pontzen et al., 2008; Tescari et al., 2009; Fumagalli et al., 2011; Cen, 2012; van de Voort et al., 2012a). To maximize the numerical resolution required for accurate modelling of the high H I column densities, most previous studies have used small simulation boxes or zoomed simulations. Those studies often try to compensate for the lack of a full cosmological distribution of absorbers by combining the results from their small-scale simulations with analytic halo mass functions to predict the properties of the DLA population (e.g., Gardner et al., 1997, 2001) or to determine what kinds of galaxies dominate the cosmic DLA distribution (e.g., Pontzen et al., 2008). This approach requires some preconceptions about the types of environments that can give rise to DLA absorbers and cannot easily account for the large scatter in the distribution of absorbers in halos with similar properties. Moreover, zoom simulations cannot fully capture the possibility of absorber-galaxy pairs being found outside of the considered regions. As a result, the statistical properties found using zoomed simulations may be biased. Finally, the impact of finite detection thresholds on the observed relation between strong H I absorbers and galaxies cannot be studied with simulations that do not contain a representative sample of H I absorbers and galaxies.

In this work, we use cosmological hydrodynamical simulations that contain a representative sample of the full distribution of strong H I systems (Rahmati et al., 2013a). Similar to what is done observationally, we connect each absorber to its nearest galaxy. A significant improvement in this work is the use of photoionization corrections that are based on accurate radiative transfer simulations and that account for both the uniform ultraviolet background (UVB) radiation and recombination radiation. In addition we show that our main conclusions are insensitive to the inclusion of local sources and to variations in the subgrid physics. The ionization corrections we use have been shown to reproduce the observed H I column density distribution function over a wide range of redshifts (Rahmati et al., 2013a).

We predict correlations between the column density of strong H I absorbers, their impact parameters, and the properties of the associated galaxies. While the fraction of H I absorbers that are linked to relatively massive galaxies increases with H I column density, most LLS and DLAs are closely associated with very low-mass galaxies, with typically  $M_{\star} \lesssim 10^8 M_{\odot}$ , that are generally undetectable with current instruments. We show that our predictions are nevertheless in good agreement with existing observations, including those of Rudie et al. (2012) who found that a large fraction of strong H I absorbers at  $z \sim 3$  are within 300 proper

kpc radius from massive Lyman-Break galaxies.

The structure of this chapter is as follows. In §4.2 we discuss our numerical simulations and ionization calculations for obtaining the HI column densities and describe our method for connecting HI systems to their host galaxies. We present our results in §4.3 and compare them with observations. In this section we also investigate how the distribution of HI absorbers varies with the properties of their host galaxies. Finally, we conclude in §4.4.

## 4.2 Simulation techniques

In this section we describe different parts of our simulations. We briefly explain the details of the hydrodynamical simulations that are post-processed to get the HI distribution by accounting for various ionization processes. Then we explain our halo finding method, our HI column density calculations, and the procedure we use to connect HI absorbers to their host halos.

### 4.2.1 Hydrodynamical simulations

We use cosmological simulations performed using a significantly modified and extended version of the smoothed particle hydrodynamics (SPH) code GADGET-3 (last described in Springel, 2005). The simulations are part of the Overwhelmingly Large Simulations (OWLS) described in Schaye et al. (2010). For our reference model, we use a subgrid pressure-dependent star formation prescription of Schaye & Dalla Vecchia (2008) which reproduces the observed Kennicutt-Schmidt law. The chemodynamics is based on the model of Wiersma et al. (2009b) which follows the abundances of eleven elements assuming a Chabrier (2003) IMF. These abundances are used for calculating radiative cooling/heating rates, element-by-element and in the presence of the uniform cosmic microwave background and the Haardt & Madau (2001) UVB model (Wiersma et al., 2009a). Galactic winds driven by star formation are modeled using a kinetic feedback recipe that assumes 40% of the kinetic energy generated by Type II SNe is injected as outflows with initial velocity of  $600 \text{ km s}^{-1}$  and with a mass loading factor  $\eta = 2$  (Dalla Vecchia & Schaye, 2008). To bracket the impact of feedback, we also consider simulations with different feedback and sub-grid models. We found that our results are not sensitive to the variations in feedback and sub-grid physics (see Appendix B).

We adopt cosmological parameters consistent with the WMAP year 7 results:  $\{\Omega_m = 0.272, \Omega_b = 0.0455, \Omega_\Lambda = 0.728, \sigma_8 = 0.81, n_s = 0.967, h = 0.704\}$  (Komatsu et al., 2011). Our reference simulation has a periodic box of  $L = 25$  comoving  $h^{-1}\text{Mpc}$  and contains  $512^3$  dark matter particles with mass  $6.3 \times 10^6 h^{-1}M_\odot$  and an equal number of baryons with initial mass  $1.4 \times 10^6 h^{-1}M_\odot$ . The Plummer equivalent gravitational softening length is set to  $\epsilon_{\text{com}} = 1.95 h^{-1}\text{kpc}$  and is limited to a minimum physical scale of

$\epsilon_{\text{prop}} = 0.5 h^{-1} \text{kpc}$ . In addition to our reference simulation explained above, we use simulations with different resolutions and box-sizes to investigate numerical effects (see Appendix D).

## 4.2.2 Finding galaxies

For identifying individual galaxies in our cosmological simulations, we assume that galaxies are bound to their dark matter haloes. This assumption implies that any given baryonic particle belongs to its closest dark matter halo. We use the Friends-of-Friends (FoF) algorithm to identify groups of dark matter particles that are near each other (i.e., FoF haloes), using a linking length of  $b = 0.2$ . Then, we use SUBFIND (Dolag et al., 2009) to connect gravitationally bound particles as part of unique structures (halos) and to identify the center of each halo/galaxy as the position of the most bound particle in that halo. We take the radius within which the average density of a given halo reaches 200 times the mean density of the Universe at a given redshift,  $R_{200}$ , as the size of that halo. The galaxy that sits in the center of each halo is considered as a *central* galaxy and all the other gravitationally bound structures in that FoF halo are considered as *satellite* galaxies. Note that we do not require satellite galaxies to be within the  $R_{200}$  of their central galaxy.

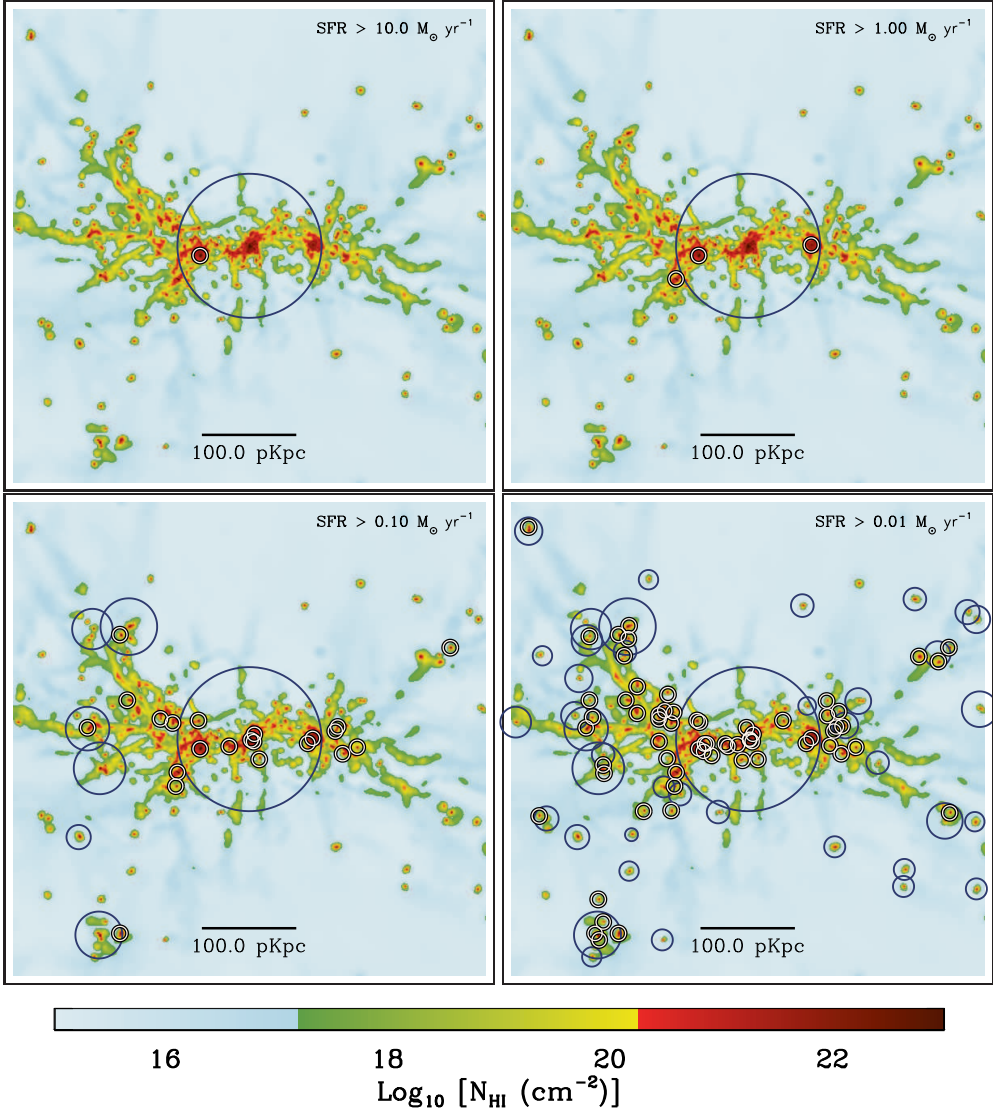
In our analysis, we use all the simulated galaxies that have star formation rates  $\text{SFR} > 4 \times 10^{-3} M_{\odot} \text{yr}^{-1}$ . By using this SFR threshold, more than 99% of our selected galaxies are resolved with  $> 100$  resolution elements (i.e., dark matter particles and/or baryonic particles). We test the impact of different SFR thresholds on our results, which provides useful insights for observational studies with finite detection threshold (see §4.3).

## 4.2.3 Finding strong HI absorbers

The first step in identifying HI absorbers in the simulations is to accurately calculate the hydrogen neutral fractions. To accomplish this, the main ionization processes that shape the distribution of neutral hydrogen must be accounted for. In this context, photoionization by the metagalactic UVB radiation is the main contributor to hydrogen ionization at  $z \gtrsim 1$  while collisional ionization becomes more important at lower redshifts (Rahmati et al., 2013a). Although the photoionization from local stellar radiation is the dominant source of ionization at high HI column densities (e.g., Rahmati et al., 2013b), our tests show that it does not have a significant impact on the results we present in this work (see Appendix C).

We use the UVB model of Haardt & Madau (2001) to account for the large-scale photoionization effect of quasars and galaxies. The same UVB model is used for calculating heating/cooling in our hydrodynamical simulations. It has been shown that this UVB model is consistent with metal absorption lines at





**Figure 4.1:** The simulated H I column density distribution around a massive galaxy with  $M_{\star} = 10^{10} M_{\odot}$  and  $\text{SFR} = 29 M_{\odot} \text{ yr}^{-1}$  at  $z = 3$ . In each panel, galaxies with different SFRs are at the center of circles. The size of dark circles indicates the virial radius of the central galaxies ( $R_{200}$ ) while the small white circles show satellite galaxies. From top-left to bottom-right, panels show galaxies with  $\text{SFR} > 10 M_{\odot} \text{ yr}^{-1}$ ,  $\text{SFR} > 1 M_{\odot} \text{ yr}^{-1}$ ,  $\text{SFR} > 0.1 M_{\odot} \text{ yr}^{-1}$  and  $\text{SFR} > 0.01 M_{\odot} \text{ yr}^{-1}$ , respectively. As the SFR threshold decreases, more galaxies are detected and the typical impact parameter between galaxies and absorbers decreases.



$z \sim 3$  (Aguirre et al., 2008) and the observed HI column density distribution function and its evolution in a wide range of redshifts (Rahmati et al., 2013a).

We characterize the UVB by its optically thin hydrogen photoionization rate,  $\Gamma_{\text{UVB}}$ , and the effective hydrogen absorption cross-section,  $\bar{\sigma}_{\text{vHI}}$  (see equations 3 and 4 and Table 2 in Rahmati et al., 2013a). In self-shielded regions,  $\Gamma_{\text{UVB}}$  is attenuated to an effective total photoionization rate,  $\Gamma_{\text{Phot}}$ , which is decreasing with density. In Rahmati et al. (2013a) we performed radiative transfer simulations of the UVB and recombination radiation in cosmological density fields using TRAPHIC (Pawlik & Schaye, 2008, 2011). We showed that the effective photoionization rate at all densities can be accurately reproduced by the following fitting function:

$$\frac{\Gamma_{\text{Phot}}}{\Gamma_{\text{UVB}}} = 0.98 \left[ 1 + \left( \frac{n_{\text{H}}}{n_{\text{H,SSH}}} \right)^{1.64} \right]^{-2.28} + 0.02 \left[ 1 + \frac{n_{\text{H}}}{n_{\text{H,SSH}}} \right]^{-0.84}, \quad (4.1)$$

where  $n_{\text{H}}$  is the hydrogen number density and  $n_{\text{H,SSH}}$  is the self-shielding density threshold given by

$$n_{\text{H,SSH}} = 6.73 \times 10^{-3} \text{ cm}^{-3} \left( \frac{\bar{\sigma}_{\text{vHI}}}{2.49 \times 10^{-18} \text{ cm}^2} \right)^{-2/3} \left( \frac{\Gamma_{\text{UVB}}}{10^{-12} \text{ s}^{-1}} \right)^{2/3}. \quad (4.2)$$

We use the photoionization rate from equations (4.1) and (4.2) together with the hydrogen number density and temperature of each SPH particle in our hydrodynamical simulations to calculate the equilibrium hydrogen neutral fraction of that particle in post-processing (see Appendix A2 in Rahmati et al., 2013a). It is also worth noting that in our hydrodynamical simulations, ISM gas particles (which all have densities  $n_{\text{H}} > 0.1 \text{ cm}^{-3}$ ) follow a polytropic equation of state that defines their temperatures. Since these temperatures are not physical and only measure the imposed pressure (Schaye & Dalla Vecchia, 2008), in our calculations we set the temperature of the ISM particles to  $T_{\text{ISM}} = 10^4 \text{ K}$ , the typical temperature of the warm-neutral phase of the ISM.

At very high HI column densities, where the gas density and the optical depth for H<sub>2</sub>-dissociating radiation is high, hydrogen is expected to be mainly molecular. This process has been suggested as an explanation for the observed cut-off in the abundance of absorbers at  $N_{\text{HI}} \gtrsim 10^{22} \text{ cm}^{-2}$  (Schaye, 2001c; Krumholz et al., 2009; Prochaska & Wolfe, 2009). It has been also shown that accounting for H<sub>2</sub> formation can produce a good agreement between cosmological simulations and observations of the HI column density distribution function (Altay et al., 2011; Rahmati et al., 2013a). To test the impact of H<sub>2</sub> formation on the spatial distribution of HI absorbers, we adopted the observationally inferred pressure law of Blitz & Rosolowsky (2006) to compute the H<sub>2</sub> fractions in post-processing (see Appendix A in Rahmati et al., 2013b). Once the H<sub>2</sub> fractions have been calculated, we exclude the molecular hydrogen from the total neutral gas for calculating the HI column densities. We note that the adopted empirical relation for calculating the H<sub>2</sub> fractions is calibrated based on observation of

low redshift galaxies and may not be accurate in very low metallicity regimes. However, due to the tight mass-metallicity-SFR relation observed out to  $z \gtrsim 3$  (Mannucci et al., 2010; Lara-López et al., 2010), exceptionally low metallicities are not very likely to be typical at  $z = 3$ , the redshift on which we focus here.

We calculate HI column densities by projecting the HI content of the simulation box along each axis onto a grid with  $10000^2$  pixels<sup>2</sup>, using SPH interpolation. While the projection may merge distinct systems along the line of sight, this is not expected to affect very high HI column density systems that are rare in the relatively small simulation boxes that we use. We tested the impact of projection effects by performing projections using multiple slices instead of the full box. In fact, our numerical experiments show that at  $z = 3$  and for simulations with box sizes comparable to that of our simulation, the effect of projection starts to appear only at  $N_{\text{HI}} < 10^{16} \text{ cm}^{-3}$ . Since the focus of our study is to characterize the properties of strong HI absorbers with  $N_{\text{HI}} \gtrsim 10^{17} \text{ cm}^{-3}$ , our results are not sensitive to the above mentioned projection effect. The top-left panel of Figure 4.1 shows an example for the distribution of the HI column density around a galaxy with stellar mass of  $M_{\star} = 10^{10} M_{\odot}$  in our simulation at  $z = 3$ .

In addition to HI column densities, we calculate the HI-weighted velocity along each line of sight (LOS),  $\langle V_{\text{LOS}} \rangle_{\text{HI}}$  and use it to constrain the position of the strongest absorber along the projection direction. That this is the same procedure used in observations to associate absorbers and galaxies. Compared to the actual LOS velocity of absorbers, calculating the HI-weighted velocity is less expensive but using it as a proxy for the position of the strongest absorber along each line of sight might be prone to projection effects. However, we found that both the  $\langle V_{\text{LOS}} \rangle_{\text{HI}}$  velocity and the actual velocity of absorber along the LOS produce nearly identical results. For calculating  $\langle V_{\text{LOS}} \rangle_{\text{HI}}$  we use the combination of local peculiar velocity of each SPH particle,  $\vec{V}_{\text{peculiar}}$ , and its Hubble velocity,  $\vec{V}_{\text{Hubble}}$ :

$$\langle V_{\text{LOS}} \rangle_{\text{HI}} = \left\langle \left( \vec{V}_{\text{peculiar}} + \vec{V}_{\text{Hubble}} \right) \cdot \hat{x}_{\text{LOS}} \right\rangle_{\text{HI}}, \quad (4.3)$$

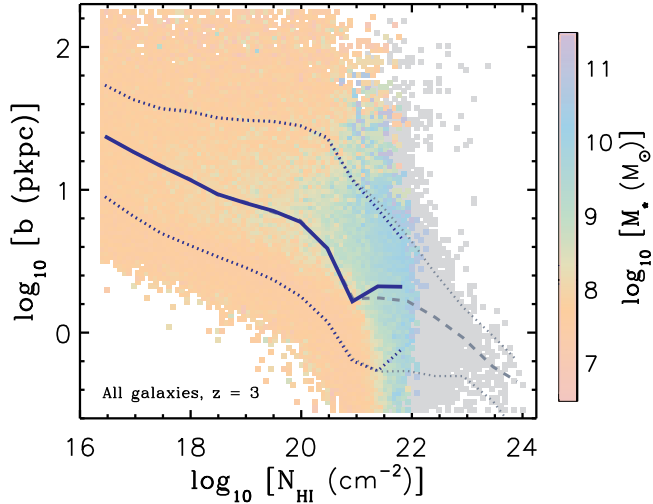
where  $\hat{x}_{\text{LOS}}$  is the unit vector along the projection direction. The Hubble velocity of each SPH particle is given by the position vector of that particle,  $\vec{X}$ , and the Hubble parameter,  $H(z)$ ,

$$\vec{V}_{\text{Hubble}} = H(z) \vec{X}. \quad (4.4)$$

We take the Cartesian coordinates of the simulation box as the reference frame. As we describe in the next section, we use the same reference frame to calculate the velocities of galaxies along the projection direction. The choice of the origin is not important since we use the velocity differences between HI absorbers and galaxies for our analysis and not the absolute velocities.

---

<sup>2</sup>Using  $10000^2$  cells produces converged results. The corresponding cell size is similar to the minimum smoothing length of SPH particles at  $z = 3$  in our simulation.



**Figure 4.2:** The predicted impact parameter of absorbers (in proper kpc) as a function of HI column density at  $z = 3$ . The color of each cell (in the 2D grid) shows the median stellar mass of the galaxies associated with the HI absorbers in that cell. The median impact parameter as a function of  $N_{\text{HI}}$  is shown using blue solid curve while the dotted curves indicate the 15% – 85% percentiles. The gray cells show the region where the  $\text{H}_2$  formation drains the atomic gas. The gray dashed and dotted curves show the median impact parameter and the 15% – 85% percentiles as a function of  $N_{\text{HI}}$  for a fully atomic HI where the conversion of high pressure gas into  $\text{H}_2$  is neglected.

#### 4.2.4 Connecting HI absorbers to galaxies

An example of how galaxies and HI absorbers are distributed in our simulation is shown in Figure 4.1. The colored map, which is repeated in all four panels, shows the HI column density distribution in a  $500^2$  proper  $\text{kpc}^2$  region which is centered on a galaxy with  $M_\star = 10^{10} M_\odot$ . Galaxies are shown with circles while the star formation rate cut for illustrating galaxies is decreasing from  $\text{SFR} > 10 M_\odot \text{yr}^{-1}$  in the top-left panel to  $\text{SFR} > 0.01 M_\odot \text{yr}^{-1}$  in the bottom-right panel. The dark circles, with sizes proportional to the virial radius of galaxies, are centered on the central galaxies and the white small circles show satellite galaxies. Galaxies with different SFRs are shown in the bottom panels of Figure 4.1. As this figure shows, galaxies and LLSs and DLAs (that are shown using green and red colors, respectively) are strongly correlated. In addition, it seems that the HI column density of absorbers is increasing as they get closer to the center of the galaxies. For a quantitative study, a well defined connection between absorbers and galaxies must be established. This connection can be made in two ways: by linking any given absorber to its closest galaxy (i.e., absorber-centered) or by finding absorbers that are closest to a given galaxy (i.e.,

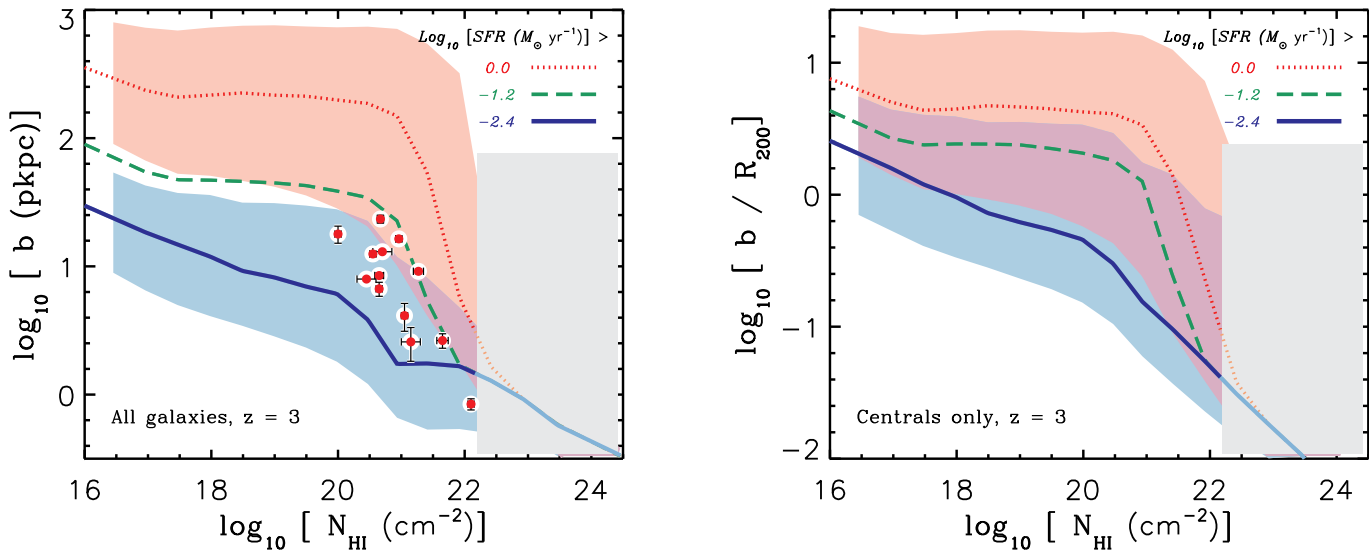
galaxy-centered). In the present work, we use the absorber-centered matching to connect the simulated HI absorbers to their neighboring galaxies. This approach is particularly efficient for associating rare strong HI absorbers to galaxies. The galaxy-centered approach, on the other hand, is more suited for studying the properties of absorbers around certain classes of galaxies. As we discuss in §4.3.6, these two approaches are closely related, but not identical.

The projected distances between HI absorbers and galaxies, together with their LOS velocity differences, can be used to associate them with each other. We use this method for a direct comparison between simulations and observational studies that employ the same approach. First, we calculate the velocity of each simulated galaxy along the LOS by adding its peculiar and Hubble velocities along the projection direction (see equation 4.3). Then, for every simulated absorber we define a galaxy counterpart that has the shortest projected distance (i.e., the impact parameter) among galaxies with the LOS velocity differences less than a chosen maximum value,  $\Delta V_{\text{LOS, max}}$ , with respect to the LOS velocity of the absorber,  $\langle V_{\text{LOS}} \rangle_{\text{HI}}$ . With this approach, each galaxy can be connected to more than one absorber, but each absorber is connected to one and only one galaxy.

We note that the difference between the LOS velocities of absorbers and galaxies includes not only the distance between the absorbers and galaxies along the LOS, but also their relative peculiar velocities along the LOS. Therefore, choosing values of  $\Delta V_{\text{LOS, max}}$  that are less than the expected peculiar velocities around galaxies results in unphysical associations between HI absorbers and neighboring galaxies. We know that accretion of the gas into halos together with galactic outflows produces typical peculiar velocities of a few hundreds of kilometers per second. Similar velocity differences have been observed between the LOS velocity of absorbers and their host galaxies (e.g., Fynbo et al., 1999; Rakic et al., 2012; Rudie et al., 2012) in addition to being common in our simulations (van de Voort & Schaye, 2012b). For this reason, we chose  $\Delta V_{\text{LOS, max}} = 300 \text{ km s}^{-1}$ , which is consistent with recent observations (Rudie et al., 2012). However, as we show in Appendix A, our results are not sensitive to this particular choice, for  $\Delta V_{\text{LOS, max}} \gtrsim 100 \text{ km s}^{-1}$ .

### 4.3 Results and discussion

Using the procedure described in the previous section, we match HI column density systems that have  $N_{\text{HI}} > 3 \times 10^{16} \text{ cm}^{-2}$  to the galaxies with non-zero SFRs in our simulations (i.e.,  $\approx 2 \times 10^6$  strong HI absorbers and more than 10000 galaxies for every projection). In the following we use this matching to study the relative spatial distribution of galaxies and HI absorbers around them.



**Figure 4.3:** Left: Predicted median impact parameters vs.  $N_{\text{HI}}$  for different SFR thresholds at  $z = 3$ . Right: Median impact parameters normalized to the virial radii ( $R_{200}$ ) as a function of  $N_{\text{HI}}$ . Since satellite galaxies reside in the halo of central galaxies, they do not have defined virial radius. Therefore, the result in the right panel is based on matching HI absorbers to central galaxies only. In both panels, the red dotted, green dashed and blue solid curves show the SFR thresholds of  $1$ ,  $0.06$  and  $0.004 M_{\odot} \text{ yr}^{-1}$ , respectively. The shaded areas around the blue solid and red dotted curves show the 15% – 85% percentiles. Red data points in the left panel show a compilation of DLAs with observed galaxy counterparts described in Table 4.1. Because of a very efficient conversion of hydrogen atoms into molecules, absorbers with  $N_{\text{HI}} \gtrsim 10^{22} \text{ cm}^{-2}$  (indicated by the gray areas) are not expected to exist.

### 4.3.1 Spatial distribution of HI absorbers

After connecting absorbers and galaxies, one can measure the typical projected distances (i.e., impact parameters,  $b$ ) separating them. The predicted distribution of impact parameters as a function of HI column density is shown in Figure 4.2 for our simulation at  $z = 3$ . Each cell in this figure shows the position of HI absorbers in the  $b - N_{\text{HI}}$  space. The color of each cell indicates the median stellar mass of galaxies that are associated with the absorbers in that cell (see the colorbar on the right-hand side). The distribution of cells only shows the range of the  $b - N_{\text{HI}}$  space that is spanned by the HI absorbers in our simulation at  $z = 3$ . To show how the impact parameter of absorbers is distributed at any given  $N_{\text{HI}}$ , we plot the median impact parameter as a function of  $N_{\text{HI}}$  using the blue solid curve and the 15% – 85% percentiles using the blue dotted curves. This result shows that our simulation predicts a strong anti-correlation between the HI column density of absorbers and their impact parameters. While the weak Lyman Limit Systems (LLSs) with  $N_{\text{HI}} \approx 10^{17} \text{ cm}^{-2}$  have typical impact parameters  $b \approx 30$  proper kpc, the impact parameter decreases with increasing the HI column density such that strong DLAs with  $N_{\text{HI}} > 10^{21} \text{ cm}^{-2}$  are typically within a few proper kpc from the center of their neighboring galaxies. The increase in the impact parameter of HI absorbers with decreasing HI column density is in agreement with observations (Moller & Warren, 1998; Christensen et al., 2007; Monier et al., 2009; Rao et al., 2011; Péroux et al., 2011; Krogager et al., 2012) and consistent with previous theoretical studies (Gardner et al., 2001; Pontzen et al., 2008).

Despite the strong anti-correlation between the impact parameter and  $N_{\text{HI}}$ , there is a large scatter around the median impact parameter at any given HI column density, as dotted curves in Figure 4.2 show. Since galaxies actively exchange material with their surroundings through accretion and outflows, the HI distribution around them has a very complex geometry (see the top-left panel of Figure 4.1 for an example). This complexity is a major contributor to the scatter in the impact parameters. In addition, part of this scatter is due to the fact that, at any given  $N_{\text{HI}}$ , there is a large number of host galaxies with different sizes that contribute to the total distribution of absorbers. This is also consistent with the color gradients in Figure 4.2: at any given impact parameter, the mass of galaxies that are linked to HI absorber is increasing with their  $N_{\text{HI}}$  and at any given HI column density (particularly for DLAs), the mass of galaxy counterparts is increasing with the impact parameter of the HI absorbers. We will discuss this further in §4.3.5.

To show the impact of  $\text{H}_2$  formation on the distribution of HI absorbers, in Figure 4.2 we also show the regions where the HI gas is fully converted into molecules using gray cells (see §4.2.3 for the details of  $\text{H}_2$  calculation). The median impact parameters and 15% – 85% percentiles for the fully atomic gas (i.e., no HI to  $\text{H}_2$  conversion) are shown with the gray dashed and gray dotted curves, respectively. The comparison between the colored and gray areas (and

curves) in Figure 4.2 shows that  $\text{H}_2$  formation only affects HI column densities  $N_{\text{HI}} > 10^{22} \text{ cm}^{-2}$ . This is consistent with the sharp cut-off in the observed HI column density distribution at  $N_{\text{HI}} > 10^{22} \text{ cm}^{-2}$  as shown in Rahmati et al. (2013a) (see also Altay et al., 2011; Erkal et al., 2012). The formation of  $\text{H}_2$  thus only drains the atomic gas at very high HI column densities and does not significantly affect the impact parameters of the HI absorbers with  $N_{\text{HI}} < 10^{22} \text{ cm}^{-2}$ .

### 4.3.2 The effect of a finite detection threshold

As seen from the colors in Figure 4.2, our simulation predicts that most strong HI absorbers with  $10^{17} < N_{\text{HI}} \lesssim 10^{21} \text{ cm}^{-2}$  are closely associated with low-mass galaxies, with typical stellar masses of  $M_{\star} \sim 10^8 M_{\odot}$ . The typical SFR for those galaxies is  $\sim 10^{-1} M_{\odot} \text{ yr}^{-1}$ . On the other hand, the typically accessible sensitivity of observations only allows the detection of galaxies that have  $\text{SFR} \gtrsim 1 - 10 M_{\odot} \text{ yr}^{-1}$  (at  $z \approx 3$ )<sup>3</sup>. Because of this relatively high detection threshold, most galaxy counterparts are not detectable and the chance of observing galaxies that host LLSs and DLAs is slim. In fact, this could be the main reason why observational surveys that are aiming to find galaxies close to DLAs, often result in non-detections (e.g., Foltz et al., 1986; Smith et al., 1989; Lowenthal et al., 1995; Bunker et al., 1999; Prochaska et al., 2002; Kulkarni et al., 2006; Rahmani et al., 2010; Bouché et al., 2012). Moreover, the finite and relatively low sensitivity of observational surveys might cause biases in the measured typical impact parameter of absorbers by mis-associating them to the closest detectable galaxy in their vicinity, instead of their real hosts that are likely to fall below the detection limit.

Galaxies in our simulation are resolved down to SFRs that are much lower than the typical limited detection threshold of observations. Therefore, we are able to analyze the impact of varying the detection limit on the impact parameter of strong HI absorbers. Figure 4.1 shows the distribution of the HI column densities and positions of galaxies in a simulated region of size 500 proper kpc around a randomly selected massive galaxy at  $z = 3$ . The top-left panel of Figure 4.1 shows the distribution of HI column density and galaxies that have  $\text{SFR} > 10 M_{\odot} \text{ yr}^{-1}$ . With this detection threshold, only the central galaxy (shown with the dark circle whose size is proportional to the virial radius of the central galaxy) and one of its satellites (shown with the small white circle) are detectable. Other panels in this figure show that as the SFR threshold for detecting galaxies decreases, more galaxies show up in the field, which decreases the typical impact parameter.

The impact of varying detection threshold on the impact parameter of strong HI absorbers (i.e.,  $N_{\text{HI}} \gtrsim 10^{17} \text{ cm}^{-2}$ ) is shown more quantitatively in the left panel of Figure 4.3. Different curves show the median impact parameter as a

<sup>3</sup>For a Chabrier IMF, a star formation rate of  $1 M_{\odot} \text{ yr}^{-1}$  corresponds to a Ly $\alpha$  luminosity of  $7.3 \times 10^{41} \text{ erg s}^{-1}$ , which translates into an observed flux of  $\approx 3 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$  and  $\approx 1 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$  at redshifts  $z = 2$  and  $z = 3$ , respectively.



**Table 4.1:** A compilation of the confirmed DLA-galaxy pairs from the literature. The columns from left to right show respectively: the ID of each DLA, their redshifts, the HI column densities, the impact parameters in arc seconds, the impact parameters in proper kpc, the star formation rates of the host galaxies and finally the references from which these values are extracted. We note that the star formation rate estimates are often based on Ly $\alpha$  emissions, which provides a lower limit for the SFR since this is difficult to account for the dust reddening and incomplete flux measurements. The SFR estimates that are corrected for the dust reddening are indicated with bold number.

ID	$z_{\text{DLA}}$	$\log N_{\text{HI}}$ ( $\text{cm}^{-2}$ )	$b$ (arcsec)	$b_{\text{p}}$ (pkpc)	SFR ( $M_{\odot} \text{ yr}^{-1}$ )	$M_{\star}$ ( $10^9 M_{\odot}$ )	Reference
Q2206-1958	1.92	20.65	0.99	8.44	3	-	[1]
Q0151+048A	1.93	20.45	0.93	7.93	71	-	[3] & [4]
PKS 0458-02	2.04	21.65	0.31	2.63	6	-	[2]
Q1135-0010	2.21	22.10	0.10	0.84	25	-	[6]
Q0338-0005	2.22	21.05	0.49	4.12	-	-	[5]
Q2243-60	2.33	20.67	2.28	23.37	<b>36</b>	-	[7]
Q2222-0946	2.35	20.65	0.8	6.67	<b>13</b>	2	[8] & [9]
Q0918+1636	2.58	20.96	2.0	16.38	<b>27</b>	12.6	[10] & [5] & [15]
Q0139-0824	2.67	20.70	1.6	13.01	-	-	[11]
J073149+285449	2.69	20.55	1.54	12.50	<b>12</b>	-	[12]
PKS 0528-250	2.81	21.27	1.14	9.15	17	-	[1]
2233.9+1318	3.15	20.00	2.3	17.91	20	-	[13]
Q0953+47	3.40	21.15	0.34	2.58	-	-	[14]

[1]- Möller et al. (2002); [2]- Möller et al. (2004); [3]- Moller & Warren (1998); [4]- Fynbo et al. (1999); [5]- Krogager et al. (2012); [6]- Noterdaeme et al. (2012); [7]- Bouché et al. (2012); [8]- Fynbo et al. (2010); [9]- Krogager et al. (2013); [10]- Fynbo et al. (2011); [11]- Wolfe et al. (2008); [12]- Fumagalli et al. (2010); [13]- Djorgovski et al. (1996); [14]- Prochaska et al. (2003); [15]- Fynbo et al. (2013)



function of  $N_{\text{HI}}$  assuming different SFR detection thresholds. The blue solid curve assumes a SFR detection threshold identical to that of Figure 4.2, where all galaxies with  $\text{SFR} > 4 \times 10^{-3} M_{\odot} \text{yr}^{-1}$  are considered as galaxy counterparts. The green dashed and red dotted curves, which respectively indicate SFR thresholds of  $> 6 \times 10^{-2} M_{\odot} \text{yr}^{-1}$  and  $> 1 M_{\odot} \text{yr}^{-1}$ , show the impact of increasing the SFR threshold on the resulting impact parameter distribution. The shaded areas around the blue solid and red dotted curves show the 15% – 85% percentiles, and the overlap region between the two shaded areas is shown in purple. The gray area at  $N_{\text{HI}} > 10^{22} \text{cm}^{-2}$  shows the region affected by the formation of  $\text{H}_2$ . The comparison between the three curves in the left panel of Figure 4.3 shows that the impact parameters increase with the detection threshold. Almost all absorbers that are associated to galaxies using the highest SFR threshold (i.e., red dotted curve) find fainter galaxies close to them (i.e., blue solid curve). Moreover, the anti-correlation between the impact parameter and  $N_{\text{HI}}$  is sensitive to the detection threshold. As the green dashed and red dotted curves show, for detection threshold  $\text{SFR} \gtrsim 10^{-1} M_{\odot} \text{yr}^{-1}$  the strong anti-correlation between the impact parameter and  $N_{\text{HI}}$  becomes insignificant at  $N_{\text{HI}} < 10^{20} \text{cm}^{-2}$ . The main reason for this is that most galaxy counterparts that are detectable with relatively low sensitivities (i.e., high detection thresholds) are not physically connected to the strong HI absorbers. As a result, the probability distribution of the impact parameters for those systems is controlled by the average projected distribution of the detectable galaxies. As the red dotted curve in the left panel of Figure 4.3 shows, our simulation predicts that with a detection threshold of  $\text{SFR} > 1 M_{\odot} \text{yr}^{-1}$  the typical impact parameters between strong HI absorbers and their nearest galaxies vary from several tens of kpc to a few hundred kpc. This result is in excellent agreement with the measured impact parameters between DLAs and galaxies in observational surveys that used similar detection thresholds at  $z \approx 2 - 3$  (not shown in this plot but see e.g., Teplitz et al., 1998; Mannucci et al., 1998). It is also worth noting that the anti-correlation between  $b$  and  $N_{\text{HI}}$  remains in place for absorbers with high HI column densities (i.e.,  $N_{\text{HI}} > 10^{21} \text{cm}^{-2}$ ), even for a relatively high detection threshold like  $\text{SFR} > 1 M_{\odot} \text{yr}^{-1}$ . However, the impact parameters of those systems are increasingly over-estimated as the galaxy detection threshold is increased.

We show the observed impact parameters of a compilation of confirmed DLA-galaxy pairs using red symbols with error-bars in the left panel of Figure 4.3 (see Table 4.1 for more details). Given the large scatter in the  $b - N_{\text{HI}}$  relation, there is a broad agreement between the observations and simulations. We note, however, that the observed impact parameters are generally above the blue solid curve. Before interpreting this systematic difference, it is important to keep in mind that there are many more non-detections of close galaxy-DLA pairs in the literature compared to those that are detected (e.g., Foltz et al., 1986; Smith et al., 1989; Lowenthal et al., 1995; Bunker et al., 1999; Prochaska et al., 2002; Kulkarni et al., 2006; Rahmani et al., 2010; Bouché et al., 2012). Therefore,

the existing observational sample is most likely unrepresentative and cannot be used for statistical analysis. Moreover, relatively high SFR detection thresholds used in the observations strongly bias the measured impact parameters towards larger values. In other words, it not easy to rule out the presence of fainter and therefore non-detected galaxies at smaller impact parameters, for DLAs that have observed bright galaxies around them. Finally, we note that SFRs of the observed galaxies that are associated with DLAs are typically  $\gtrsim 10 M_{\odot} \text{ yr}^{-1}$  (see Table 4.1), which implies that those galaxies are massive and therefore large systems. As we show in §4.3.5, the observed impact parameters of those DLA-galaxy pairs are consistent with the distribution of impact parameters around similar galaxies in our simulation.

### 4.3.3 Distribution of HI absorbers relative to halos

It is useful to compare the impact parameters that connect absorbers to their neighboring galaxies with the size of their halos. We therefore define the normalized impact parameter of absorbers as the ratio of the impact parameter and the virial radius ( $R_{200}$ ) of the host galaxies. Since the virial radius is well defined only for central galaxies, we only consider those objects when associating absorbers to galaxies. As shown by the blue solid curve in the right panel of Figure 4.3, strong HI absorbers tend to be located closer to the center of haloes as their  $N_{\text{HI}}$  increases. Most LLSs are found within one virial radius of their host galaxies and the majority of DLAs is found within a few tenths of viral radius from their associated central galaxies. There is, however, a large scatter in the normalized impact parameters at given  $N_{\text{HI}}$  which is shown by the shaded area around blue solid and red dotted curves in the right panel of Figure 4.3. As implied by this figure, a non-negligible fraction of DLAs are expected at impact parameters comparable to, or even larger than the virial radius of their associated central galaxies. This is, in part, due to neglecting satellite galaxies in matching absorbers and galaxies (because they do not have a well defined virial radius) which effectively associates the strong HI absorbers that are near satellites to their closest central galaxies<sup>4</sup>. For the same reason, the median impact parameter at a given  $N_{\text{HI}}$  increases by leaving the satellite galaxies out of the analysis (not shown). Moreover, the complex and highly structured distribution of strong HI absorbers which often extends to distances beyond the virial radius of central galaxies (see Figure 4.1) also contributes to the large scatter around the median normalized impact parameter at a given  $N_{\text{HI}}$ .

There is also an anti-correlation between the normalized impact parameter of absorbers and their HI column density, which is steeper than the relation we see between the absolute impact parameters and  $N_{\text{HI}}$  (shown in the right panel of Figure 4.3). This difference is most pronounced at  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$  where

---

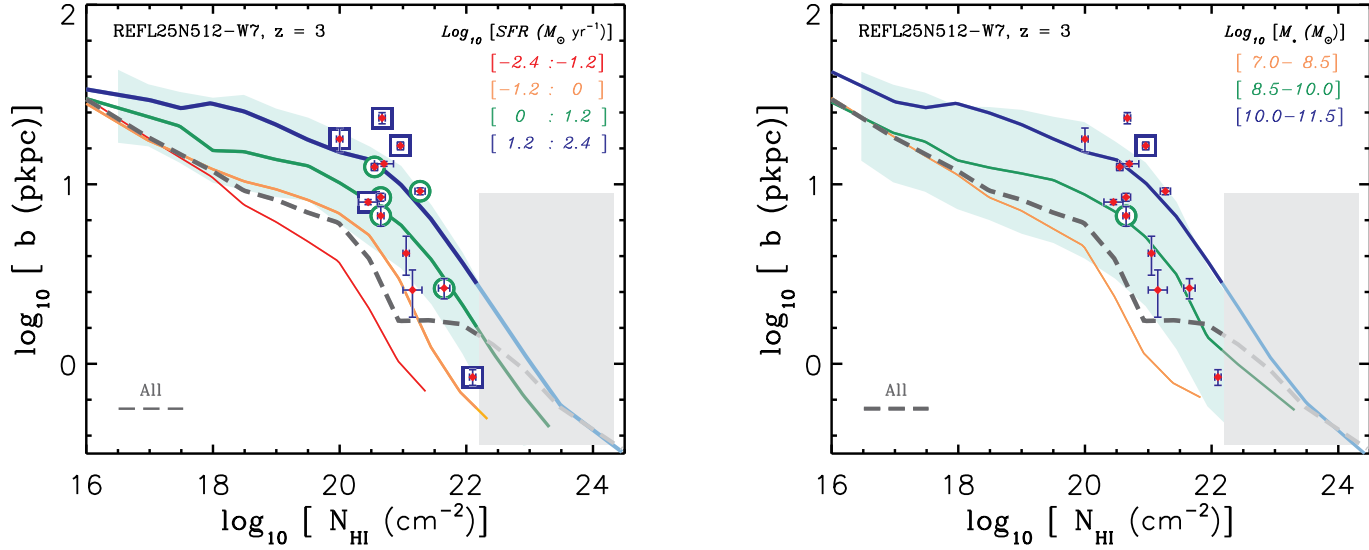
<sup>4</sup>Note that satellite galaxies (and hence their associated absorbers) are not necessarily within the virial radius of their host galaxies, due to the non-spherical distribution of FoF structures. This is shown in Figure 4.1.

the impact parameter flattens with increasing  $N_{\text{HI}}$  but the normalized impact parameters still decrease steeply with  $N_{\text{HI}}$ . This trend can be explained by the contribution of very massive (and hence large) galaxies becoming increasingly more dominant at very high HI column densities (see §4.3.5). Therefore, there are two competing trends: on the one hand, systems with higher HI column densities tend to be closer to the center of their neighboring (host) haloes. On the other hand, the size of the haloes that are linked to those absorbers becomes larger with increasing column density. These two effects neutralize each other at very high HI column densities, forming a less steep  $b - N_{\text{HI}}$  relation at  $N_{\text{HI}} > 10^{21} \text{ cm}^{-2}$  compared to that of the lower column densities. By normalizing the impact parameter of absorbers to the virial radii of their associated haloes, the latter effect (i.e., the increasing size of the dominant haloes with increasing  $N_{\text{HI}}$ ) is canceled out.

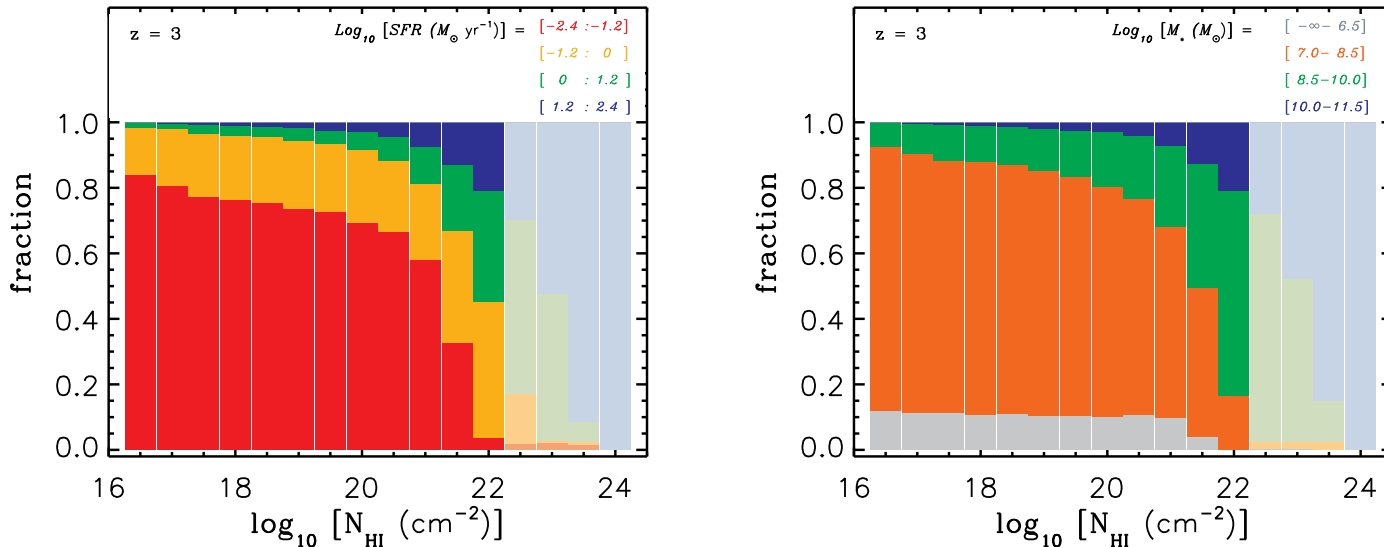
The trends we discussed earlier for the effect of varying the SFR threshold on the impact parameter of absorbers hold qualitatively for the normalized impact parameters as well. As the green dashed and red dotted curves in the right panel of Figure 4.3 show, increasing the SFR threshold for galaxies that are considered in matching absorbers to galaxies, results in larger normalized impact parameters. Despite the qualitatively similar trends, the differences between the normalized impact parameters for different SFR thresholds are smaller than the differences between absolute impact parameters. At  $10^{17} < N_{\text{HI}} < 10^{21} \text{ cm}^{-2}$ , both the green dashed and red dotted curves in the right panel of Figure 4.3 are nearly flat. This implies that absorbers at those HI column densities are several virial radii away from their closest central galaxies (with  $\text{SFR} > 6 \times 10^{-2} M_{\odot} \text{ yr}^{-1}$  and  $\text{SFR} > 1 M_{\odot} \text{ yr}^{-1}$ , respectively). This supports our earlier statement that the association between bulk of strong HI absorbers and galaxies that have relatively high SFRs is not physical (particularly at  $N_{\text{HI}} < 10^{21} \text{ cm}^{-2}$ ), if fainter galaxies be omitted.

#### 4.3.4 Resolution limit in simulations

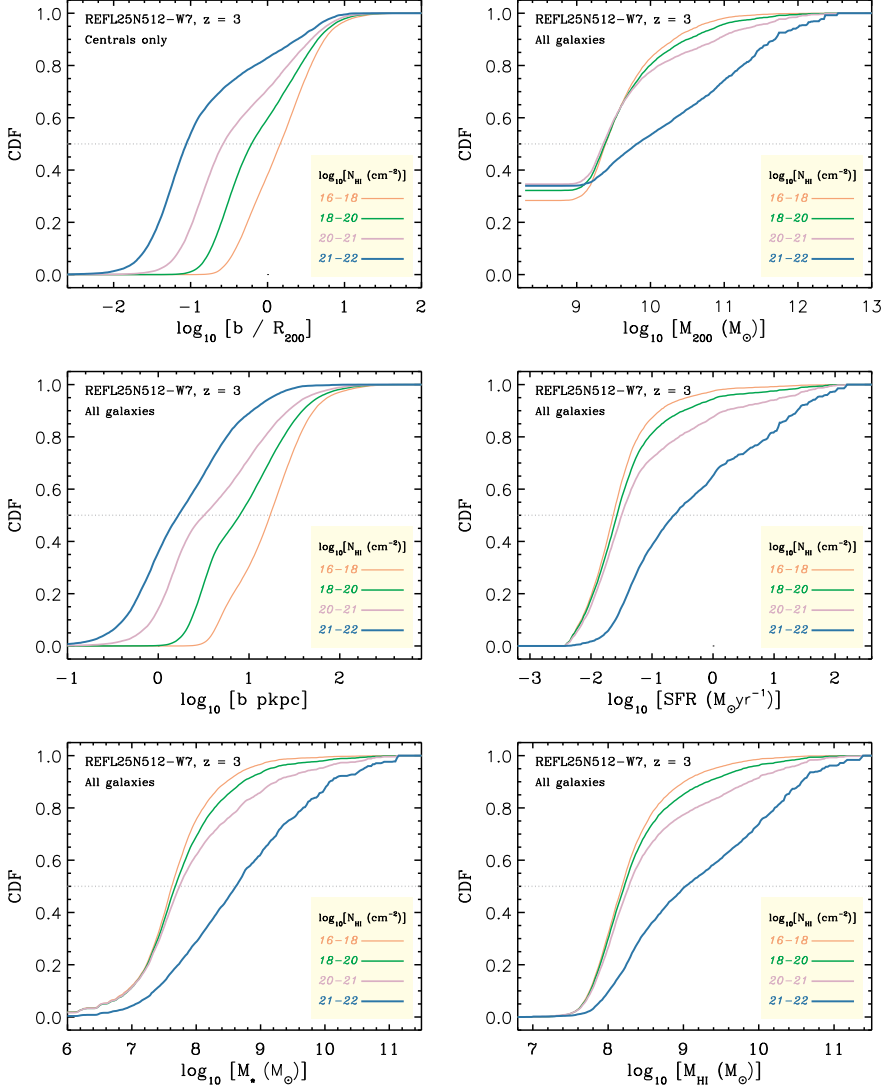
The simulated HI distribution and its connection to galaxies are both sensitive to the high resolution adopted in our simulations. The HI column density distribution function is converged for LLSs and most DLAs at the resolution we use in this work (Rahmati et al., 2013a), but this does not rule out that their distribution relative to galaxies is also resolution independent. The relative position of galaxies that are identified as bound structures in a cosmological simulation is mainly determined by the distribution of cosmological overdensities, and therefore, is not expected to be highly resolution dependent. However, by increasing the resolution, more and more low-mass galaxies start to be identified which might affect the typical distances between strong HI absorbers and galaxies. Because of this, it is important to adopt a lower limit for galaxies that are taken into account in our analysis to make sure our findings are not driven primarily by the resolution we use. At the same time, this lower limit should be such that it



**Figure 4.4:** The predicted median impact parameters for subsets of HI absorbers associated with galaxies in different star formation rate bins (left) and stellar mass bins (right) as a function of  $N_{\text{HI}}$ , at  $z = 3$ . The shaded area around the solid green curves in the left (right) panel shows the 15% – 85% percentile in the distribution of absorbers that are linked to galaxies with  $1 < \text{SFR} < 16 M_{\odot} \text{ yr}^{-1}$  ( $2 \times 10^8 < M_{\star} < 10^{10} M_{\odot}$ ). The dashed curves in both panels show the median impact parameter of all absorbers as a function of  $N_{\text{HI}}$ . The data points show the observed impact parameters for the confirmed DLA-galaxy pairs. The colored circles and squares around the data points show the SFR/mass bin to which they belong. Note that the squares and circles, which show the two bins with the highest values (of SFR or mass) respectively, are in agreement with our results. Because of a very efficient conversion of hydrogen atoms into molecules, absorbers with  $N_{\text{HI}} \geq 10^{22} \text{ cm}^{-2}$  (indicated with the gray areas) are not expected to exist.



**Figure 4.5:** The top panels show the predicted median impact parameters for subsets of HI absorbers associated with galaxies in different star formation rate bins (left) and stellar mass bins (right) as a function of  $N_{\text{HI}}$ , at  $z = 3$ . The fraction of absorbers that are associated with galaxies in different star formation rate bins (left) and stellar mass bins (right) to the total number of strong HI absorbers as a function of  $N_{\text{HI}}$ , at  $z = 3$ . The star formation rate bins and stellar mass bins are identical to that of Figure 4.4. Because of a very efficient conversion of hydrogen atoms into molecules, absorbers with  $N_{\text{HI}} \gtrsim 10^{22} \text{ cm}^{-2}$  (indicated with the gray areas) are not expected to exist.



**Figure 4.6:** The top-left panel shows the predicted cumulative distribution of normalized impact parameter of absorbers in different  $N_{\text{HI}}$  bins at  $z = 3$ , where only central galaxies are taken into account. The other five panels show the cumulative distribution of different properties where all galaxies (with  $\text{SFR} > 0.004 M_{\odot} \text{ yr}^{-1}$ ) are taken into account. The top-right panel shows the cumulative distribution of halo masses and other panels from middle-left to bottom right respectively show the cumulative distribution of impact parameters, star formation rates (of associated galaxies), stellar masses and HI masses. Except for strong DLAs with  $10^{21} < N_{\text{HI}} < 10^{22} \text{ cm}^{-2}$ , bulk of HI absorbers reside close to galaxies with similar properties (i.e., virial mass, stellar mass, gas mass and star formation rate). However, their typical (normalized) impact parameters are significantly different.

allows for the inclusion of as many galaxies as possible, without throwing away useful information.

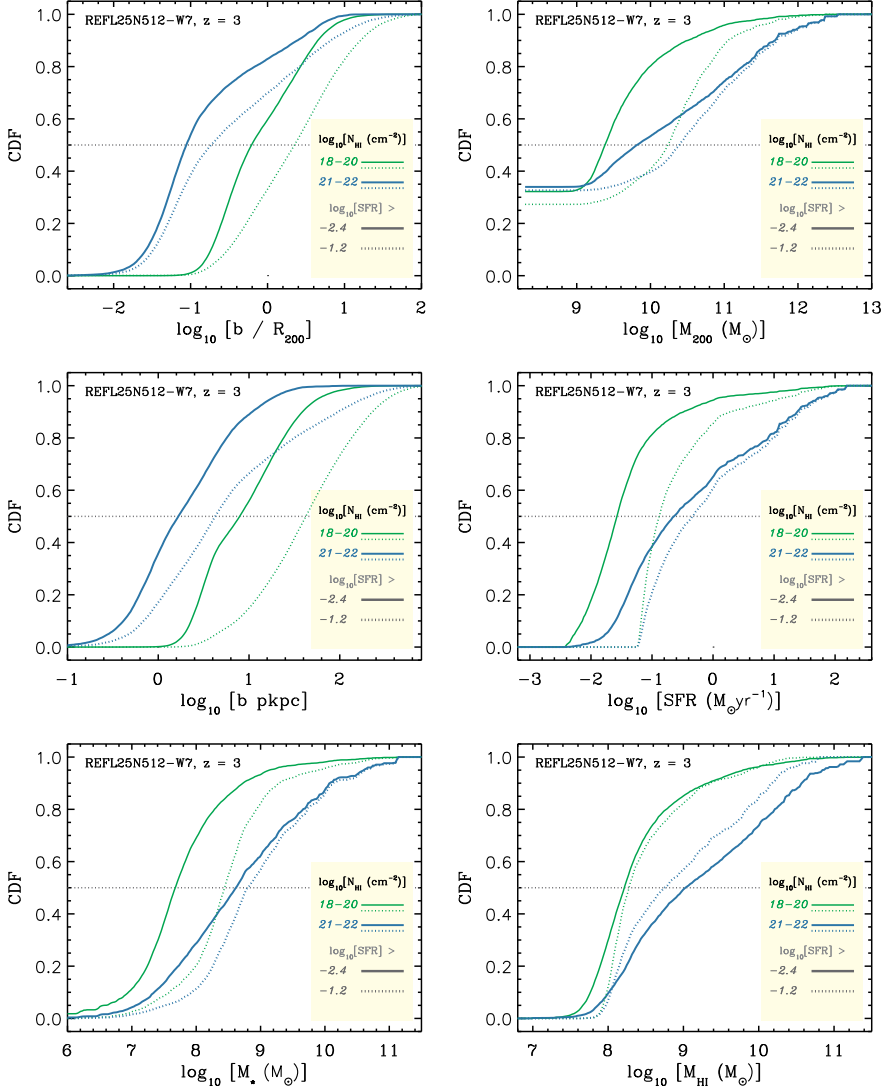
We use SFR as a quantity to limit galaxies that are considered in our analysis, which is the same quantity we used in the previous section as a proxy for the observational detectability of galaxies. The advantage of using the SFR of galaxies to identify them is the better sampling it provides compared to other reasonable quantities such as the stellar mass (see the discussion in §4.2 in Rahmati et al., 2013b). We adopt  $\text{SFR} > 4 \times 10^{-3} \text{ M}_{\odot} \text{ yr}^{-1}$  as a threshold and include only galaxies that have SFRs above this threshold in the analysis. Most bound (sub) structures (i.e.,  $\approx 98\%$ ) that are found in our simulation have SFRs above this threshold.

It is reasonable to expect that an increase in the number density of galaxies would decrease the typical distance between absorbers and galaxies. This would mean that the cumulative number density of galaxies in the simulation is the key factor that changes the typical distances between strong H<sub>I</sub> absorbers and galaxies. We show in Appendix D that this is indeed the case, and the  $b - N_{\text{HI}}$  relation is converged with resolution if the cumulative number density of galaxies that are considered is the same. Unlike the position of galaxies, the lowest SFRs in simulations are sensitive to the resolution and tend to decrease with increasing the resolution for a fixed number density. As a result, we do not expect that our result to remain the same if the resolution of our simulation increases for the same adopted SFR threshold (i.e.,  $\text{SFR} > 4 \times 10^{-3} \text{ M}_{\odot} \text{ yr}^{-1}$ ). Instead, we expect our results to be insensitive to the resolution if the adopted threshold be such that the cumulative number density of galaxies be identical to the corresponding value in our simulation. We note that the number density of galaxies in our simulation that satisfy the adopted criterion is 0.5 galaxy per comoving  $\text{Mpc}^3$  (i.e., equivalent to 31.5 galaxy per proper  $\text{Mpc}^3$ ).

Nonetheless, we found that the  $b - N_{\text{HI}}$  relation for various SFR threshold is converged if  $\text{SFR} > 6 \times 10^{-2} \text{ M}_{\odot} \text{ yr}^{-1}$ . In other words, the total number density of galaxies with  $\text{SFR} > 6 \times 10^{-2} \text{ M}_{\odot} \text{ yr}^{-1}$  is not expected to change by increasing the resolution of our simulation. This is another way of stating that for galaxies that satisfy the above mentioned condition, the SFR is converged with respect to the resolution.

### 4.3.5 Correlations between absorbers and various properties of their associated galaxies

The gas content of galaxies is correlated with their other properties like stellar mass, size and star formation rate. This implies also correlations between the abundance and distribution of strong H<sub>I</sub> absorbers and properties of galaxies that are associated with them. For instance, as mentioned earlier, at a given impact parameter, the typical stellar mass of host galaxies increases with increasing H<sub>I</sub> column density. Similarly, at a fixed  $N_{\text{HI}}$ , the typical host stellar mass increases with increasing the impact parameter (see Figure 4.2). Moreover,



**Figure 4.7:** The impact of SFR threshold on the cumulative distributions that are shown in Figure 4.6. In each panel the blue and green solid curves show the cumulative distribution of LLSs (i.e.,  $10^{18} < N_{\text{HI}} < 10^{20} \text{ cm}^{-2}$ ) and strong DLAs (i.e.,  $10^{21} < N_{\text{HI}} < 10^{22} \text{ cm}^{-2}$ ), respectively. While the solid curves, that are identical to those shown in Figure 4.6, indicate the SFR threshold of  $4 \times 10^{-3} M_{\odot} \text{ yr}^{-1}$ , the dotted curves show the result obtained by imposing a SFR threshold of  $6 \times 10^{-2} M_{\odot} \text{ yr}^{-1}$ . Panels from top-left to bottom-right show the cumulative distribution of normalized impact parameters, halo masses, impact parameters, SFRs, stellar masses and H I masses, respectively. The detection of galaxies with SFRs as low as  $\text{SFR} \sim 10^{-2} M_{\odot} \text{ yr}^{-1}$  future instruments such as MUSE.



the LLSs are closely linked to low-mass galaxies with typical stellar masses of  $M_\star \sim 10^8 M_\odot$ , in contrast to strong DLAs which tend to be associated with more massive galaxies.

The above mentioned trends are shown more clearly in Figure 4.4. These plots illustrate the  $b - N_{\text{HI}}$  relation for subsets of HI absorbers that are linked to galaxies with different star formation rates and stellar masses (respectively shown in the left and right panels). The colored curves in the left (right) panel which show different bins of SFR (stellar mass), increasing by going from red to blue indicate that the impact parameter of absorbers is increasing with the SFR and mass of their associated galaxies. Note that there is a relatively large scatter around the median impact parameter of each subset of absorbers, as shown with the shaded area around the green curves. This result explains the color gradient that we see in Figure 4.2 and is in good agreement with observations that are shown with the red diamonds in Figure 4.4. The observed DLA-galaxy pairs with reliable SFR or mass estimate are shown with the green circles and blue cubes, respectively for the highest two bins. Note that the curves are close to the observational results (see also Table 4.1).

We also show the median impact parameters of all absorbers using the dashed curve in Figure 4.4. The comparison between this curve and colored solid curves confirms our earlier conclusion (based on the color gradient in Figure 4.2) that the impact parameter of absorbers with higher HI column densities indicates that they are more likely to be around more massive galaxies with higher SFRs. To illustrate this more clearly, we show also the fraction of HI absorbers, at any given column density, that are associated to galaxies with a particular property. The result of this analysis for different SFR and mass bins is shown in the left and right panels of Figure 4.5, respectively. These results clearly show that the fraction of HI absorbers associated with massive galaxies (which also have high SFRs), is decreasing rapidly with decreasing their HI column densities. Also, it is clear from the left panel of Figure 4.5 that most absorbers with  $N_{\text{HI}} < 10^{21} \text{ cm}^{-2}$  are closely linked to galaxies with  $\text{SFR} < 6 \times 10^{-2} M_\odot \text{ yr}^{-1}$ , or equivalently  $M_\star < 10^8 M_\odot$ . While there are only about 20 – 30% of those systems associated with more massive galaxies with higher SFRs (i.e.,  $M_\star > 10^8 M_\odot$  and  $\text{SFR} > 6 \times 10^{-2} M_\odot \text{ yr}^{-1}$ ), the same galaxies are associated with a large fraction of strong DLAs with  $N_{\text{HI}} > 10^{21} \text{ cm}^{-2}$ . We reiterate that these results are not changing by increasing the resolution and only the lowest SFRs are expected to be reduced at higher resolutions (shown with the red regions in the left panel of Figure 4.5), which is not expected to affect the fraction of absorbers that are associated with well-resolve galaxies that have  $M_\star > 10^8 M_\odot$  and  $\text{SFR} > 6 \times 10^{-2} M_\odot \text{ yr}^{-1}$ .

We note that our results are in agreement with Tescari et al. (2009) and van de Voort et al. (2012a). Namely, those authors found that the largest fraction of HI absorbers with  $N_{\text{HI}} < 10^{21} \text{ cm}^{-2}$  are in very low-mass haloes with  $M_{200} < 10^{10} M_\odot$  (i.e.,  $M_\star \lesssim 10^8 M_\odot$ ; see the orange regions in the right panel of Figure 4.5). Moreover, van de Voort et al. (2012a) found that at higher HI

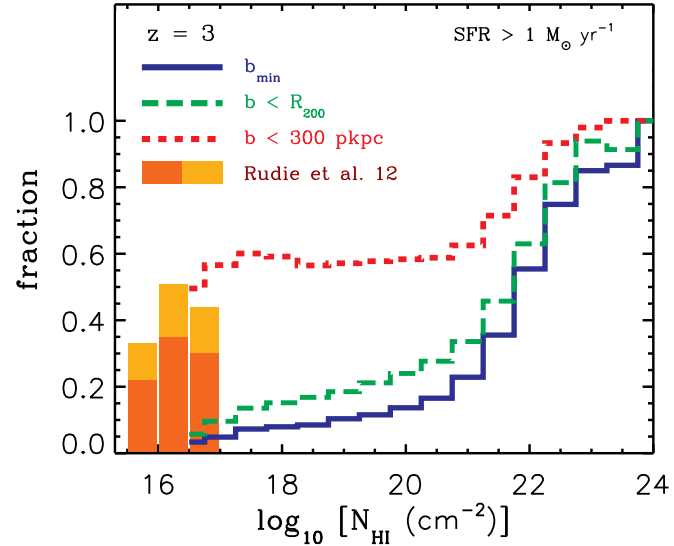
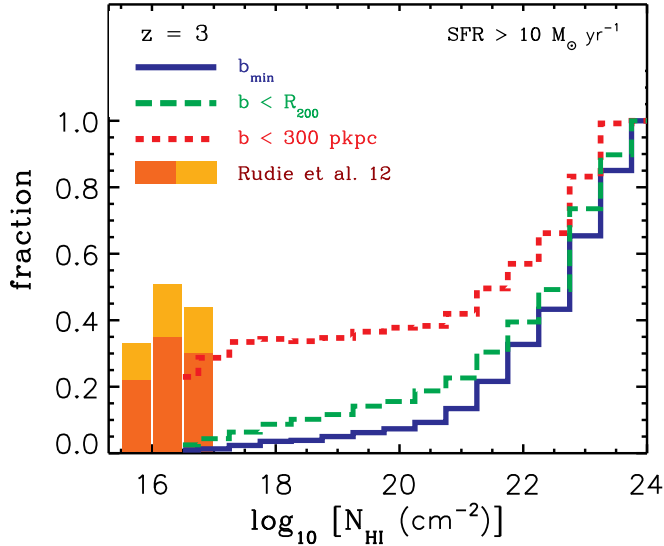
column densities, the contribution of more massive haloes with  $M_{200} > 10^{11} M_{\odot}$  (i.e.,  $M_{*} \gtrsim 10^9 M_{\odot}$ ; see the green regions in the right panel of Figure 4.5) rapidly increases.

To investigate further the distribution of galaxy properties for absorbers with different HI column densities, we split absorbers in different  $N_{\text{HI}}$  bins. Then, for galaxies that are associated with the absorbers in that bin, we construct the cumulative distribution of different properties. The result of this exercise is shown in Figure 4.6. The top-left panel of this figure shows the cumulative distribution of normalized impact parameters, where in contrary to the other five panels, only central galaxies are taken into account (because satellite galaxies do not have well defined virial radius required for the normalization). The top-right panel shows the cumulative distribution of halo masses. Since satellite galaxies are in the halo of their centrals, they do not have well defined halo masses and as the starting point of the cumulative distribution functions in the top-right panel indicates,  $\approx 30\%$  of absorbers are associated to satellite galaxies. The other four panels from middle-left to the bottom-right respectively show the cumulative distribution of impact parameters, star formation rates, stellar masses and HI masses.

Comparing the cumulative distribution of (normalized) impact parameters with other four panels in this figure indicates that the HI column density of absorbers is mostly sensitive to their projected distance from their associated galaxies (i.e., the impact parameter), compared to other properties, such as the stellar mass, SFR, HI mass or halo size of their associated galaxies. As shown in the top-left panel, while more than 50% of strong DLAs are within  $R \lesssim 0.1 R_{200}$ , most weak LLSs with  $N_{\text{HI}} \lesssim 10^{18} \text{ cm}^{-2}$  are likely to be beyond the virial radius of their host galaxies.

Another result that is shown in Figure 4.6 is that the fraction of absorbers that are linked to more massive galaxies is increasing by increasing their HI column densities. The same trend is visible for the SFRs, halo masses and HI masses. However, as the distinction between strong DLAs  $N_{\text{HI}} > 10^{22} \text{ cm}^{-2}$  (blue solid curves) and lower HI column densities show, galaxies associated with strong DLAs have distinct distributions in SFR, HI mass, stellar mass and halo mass. For instance, the median SFR of galaxies that are associated with other strong HI absorbers is  $\approx 10$  times lower than the same quantity for galaxies associated with strong DLAs. Similar trend also holds for HI masses, stellar and halo masses. Comparing the median properties (i.e., at CDF = 0.5) shows that only  $\approx 10\%$  of strong DLAs are linked to galaxies that are typically associated with weaker HI absorbers. This suggests that strong DLAs with  $N_{\text{HI}} > 10^{21} \text{ cm}^{-2}$  are preferentially linked to the most massive galaxies while other HI absorbers are distributed among more abundant galaxies with lower masses.

The middle-right panel of Figure 4.6 is particularly useful to understand the difficulty of finding the observed galaxy counterparts of strong HI absorbers. The distribution of SFRs shown in this plot can be used to predict the chance by which a galaxy counterpart can be detected above a cer-



**Figure 4.8:** Fraction of strong HI systems that are associated to galaxies with  $\text{SFR} > 10 M_{\odot} \text{ yr}^{-1}$  (left) and  $\text{SFR} > 1 M_{\odot} \text{ yr}^{-1}$  (right) in our simulation at  $z = 3$ . Curves with different line styles and colors show results that are obtained using different methods for associating HI absorbers to galaxies that are above the imposed SFR threshold: blue solid curves are obtained by associating HI absorbers to their closest galaxies, where all the simulated galaxies are taken into account (see Figures 4.4 and 4.5); green long-dashed curves show the fraction of absorbers that reside within the virial radius of the selected galaxies, and finally, red short-dashed curves show the fraction of absorbers that are within 300 proper kpc from the selected galaxies. For comparison, the colored bars show Rudie et al. (2012) findings for fraction of absorbers within 300 proper kpc from galaxies with similar SFRs that are shown here. The dark orange part of the bar is the observed fractions and the light orange part shows the correction for missing galaxies in their spectroscopic sample. Predicted fractions are in excellent agreement with observations and show a large fraction of strong HI absorbers are less than 300 proper kpc away from galaxies with  $\text{SFR} > 1 M_{\odot} \text{ yr}^{-1}$ . However, most of those systems are far beyond the virial radius of those galaxies and are associated with less massive objects with lower SFRs which are below the detection limit of observations.

tain detection threshold. For instance, if the detection threshold is equivalent to  $\text{SFR} = 1 M_{\odot} \text{ yr}^{-1}$ , the chance of detecting the galaxy counterpart of a strong DLA with  $N_{\text{HI}} \approx 10^{22} \text{ cm}^{-2}$  is 3 in 10. In other words, using this detection threshold results in non-detection rate of 70%. For weak DLAs (i.e.,  $10^{20} < N_{\text{HI}} < 10^{21} \text{ cm}^{-2}$ ) the non-detection rate would be even higher at 90%. This explains the large number of non-detections in observational studies (e.g., Foltz et al., 1986; Smith et al., 1989; Lowenthal et al., 1995; Bunker et al., 1999; Prochaska et al., 2002; Kulkarni et al., 2006; Rahmani et al., 2010; Bouché et al., 2012).

As mentioned before, with the typical detection threshold of observations at  $z \sim 3$ , only galaxies that have  $\text{SFR} > 1 - 10 M_{\odot} \text{ yr}^{-1}$  can be identified. On the other hand, galaxies are resolved to much lower SFRs in our simulation and it is not straightforward to compare current observations and the results shown in Figure 4.6. With the advent of future surveys like MUSE (Bacon et al., 2010), the accessible Ly $\alpha$  detection thresholds would be pushed to lower SFRs which allows the identification of galaxies with SFRs as low as  $\text{SFR} \sim 10^{-2} - 10^{-1} M_{\odot} \text{ yr}^{-1}$  at  $z \sim 3$ . These deep observations can be used to identify faint galaxies associated with strong HI absorbers and analyze the cumulative distribution of their properties. However, the results would still depend on the accessible detection threshold and might be different from what is shown in Figure 4.6. To address this issue, in Figure 4.7 we compare the cumulative distributions for two different detection thresholds. Panels in this figure are identical to Figure 4.6, but show the result for only two HI column density bins. The green and blue curves, respectively, represent the HI absorbers (or galaxies associated to them) that have  $10^{18} < N_{\text{HI}} < 10^{20} \text{ cm}^{-2}$  (i.e., LLSs) and  $10^{21} < N_{\text{HI}} < 10^{22} \text{ cm}^{-2}$  (i.e., strong DLAs). The solid curves show our fiducial detection threshold of  $\text{SFR} > 4 \times 10^{-3} M_{\odot} \text{ yr}^{-1}$ , which has been used in Figure 4.6. The dotted curves indicate a higher detection threshold of  $\text{SFR} > 6 \times 10^{-2} M_{\odot} \text{ yr}^{-1}$ , which is comparable to what would be accessible using deep MUSE observations. As the difference between the solid and dotted curves show, the predicted/observed distributions are sensitive to the detection threshold. In other words, the bias introduced by the limited detection threshold should be taken into account for interpreting/modeling the observed distributions.

### 4.3.6 Are most strong HI absorbers at $z \sim 3$ around Lyman-Break galaxies?

In this work, we adopted the absorber-centered point of view in which each absorber is associated with its closest galaxy. This approach is particularly efficient to establish a relationship between rare absorbers and their galaxy counterparts. As we showed in the previous sections, our simulation predicts that for most HI absorbers, nearby galaxies are low-mass objects (i.e.,  $M_{\star} \approx 10^8 M_{\odot}$ ) with low SFRs. We showed that this prediction is consistent with the impact parameters of observed DLA-galaxy counterparts, and more importantly, the high incident

rate of finding no detectable galaxy close to DLAs. There is, however, an alternative approach which is to search for HI absorbers around galaxies (i.e., galaxy-centered point of view) (e.g., Steidel et al., 1995; Adelberger et al., 2003, 2005; Hennawi & Prochaska, 2007; Steidel et al., 2010; Rakic et al., 2012; Rudie et al., 2012; Prochaska et al., 2013). While the two approaches are complementary, their results might seem inconclusive. For instance, using galaxy-centered approach, Rudie et al. (2012) found that at  $z \approx 2 - 3$  most Lyman Limit absorbers (i.e.,  $N_{\text{HI}} \approx 10^{17} \text{ cm}^{-2}$ ) are within an impact parameter of 300 proper kpc, and within 300 km/s LOS velocity difference with respect to rest-frame UV-selected star-forming galaxies (see their Figure 30). Given that the typical galaxy mass in their sample is  $M_{\star} \approx 10^{10} M_{\odot}$ , one might conclude that their result is in conflict with our finding that most strong HI absorbers (i.e., LLSs and DLAs) are closely associated with galaxies with  $M_{\star} \approx 10^8 M_{\odot}$ .

To understand the source of this discrepancy, first we note that the 300 proper kpc transverse distance and the 300 km/s LOS velocity difference that are adopted by Rudie et al. (2012), are, respectively, 300 and 50 per cent larger than the virial radius and the circular velocity which is expected for the host haloes of their galaxies (Rakic et al., 2012, 2013). In other words, the region Rudie et al. (2012) define as the “circumgalactic medium” lies well beyond the virial radius of the haloes that are thought to host their galaxies.

In addition, it is important to note that galaxies are strongly clustered and low-mass galaxies prefer to live around more massive galaxies. This implies that most strong HI absorbers are also likely to be found close to massive galaxies, while they actually belong to their low-mass parent galaxies. As a result, searching for HI absorbers within a reasonably large radius around massive, and hence easily observable galaxies, recovers a large fraction of the existing strong HI absorbers. This effect can be seen in the bottom-right panel of Figure 4.1, which shows the HI column density distribution around a galaxy with  $M_{\star} = 10^{10} M_{\odot}$ . While the maximum projected distance between the galaxy and HI absorbers that are shown around it is less than 300 proper kpc, most of those absorbers have low-mass galaxies very close to them.

The aforementioned arguments are illustrated more quantitatively in Figure 4.8, which shows the fraction of strong HI absorbers in our simulation at  $z = 3$  that are in the vicinity of galaxies with star formation rates comparable to that of LBG galaxies at the similar redshifts<sup>5</sup>. The red dashed curve in the left (right) panels shows the fraction of HI absorbers that are within the impact parameter of 300 proper kpc from galaxies with  $\text{SFR} > 10 M_{\odot} \text{ yr}^{-1}$  ( $\text{SFR} > 1 M_{\odot} \text{ yr}^{-1}$ ), as a function of  $N_{\text{HI}}$ . This result is in excellent agreement with the fractions Rudie et al. (2012) found (shown with the colored bars) and predicts that the fraction of absorbers that are within 300 proper kpc from these galaxies is roughly constant for  $10^{16} < N_{\text{HI}} < 10^{21} \text{ cm}^{-2}$ . However, only a small fraction of LLSs that are within the impact parameter of 300 proper kpc from those

<sup>5</sup>The star formation rate of galaxies used in Rudie et al. (2012) is between several to a few hundreds of  $M_{\odot} \text{ yr}^{-1}$ .

galaxies are also within their virial radii (shown with the green long-dashed curves). As the blue solid curves show, even smaller fractions of LLSs remain associated to those galaxies if we account for galaxies with lower SFRs, which are typically closer to LLSs but are too faint to be observed.

## 4.4 Summary and conclusions

We have used cosmological simulations that have been post-processed using accurate radiative transfer corrections that account for photoionization by the UVB and recombination radiation to investigate the relation between strong HI absorbers (i.e., LLSs and DLAs) and galaxies at  $z = 3$ . The simulation we used for our study has been shown to reproduce the observed HI column density distribution function accurately (Altay et al., 2011; Rahmati et al., 2013a). After identifying sight-lines with high HI column densities (i.e.,  $N_{\text{HI}} > 3 \times 10^{16} \text{ cm}^{-2}$ ) and calculating the line-of-sight velocity of absorbers, we used a procedure similar to that used in observational studies to associate absorbers with nearby galaxies. Namely, we associated each strong HI absorber to the galaxy which has the shortest transverse distance to the absorber and a line-of-sight velocity difference within  $\pm 300 \text{ km s}^{-1}$ . Associating all strong HI absorbers in the periodic simulation box to galaxies close to them allowed us to predict statistical trends between the strength of the absorbers and the distance to, and properties of, the galaxies.

Among the various dependencies we studied in this work, we found that the anti-correlation between the HI column density of absorbers and the transverse distance that connects them to their neighbouring galaxies (i.e., the impact parameter) to be the strongest. While LLSs have impact parameters  $\gtrsim 10$  proper kpc, DLAs are typically within a few proper kpc from the nearest galaxies. Relative to the virial radius of the halo that hosts the nearest central galaxy, LLSs have typical impact parameters  $\gtrsim 1 R_{200}$ , while DLAs are typically  $\sim 10$  times closer to the center of the nearest halo (i.e.  $\lesssim 0.1 R_{200}$ ). The predicted strong anti-correlation between the impact parameter of strong HI absorbers and their HI column densities is in agreement with observations and previous work. We also found a relatively large scatter around the median impact parameter of absorbers, at a given  $N_{\text{HI}}$ , due to the complex geometry of gas distribution around galaxies, and also the variation in the size and gas content of the galaxies that are producing the absorbers.

We predict that most strong HI absorbers are closely associated with relatively low-mass galaxies,  $M_{\star} \lesssim 10^8 M_{\odot}$ , but that the fraction of strong HI absorbers that are linked to more massive galaxies increases rapidly with the HI column density. This correlation between column density and galaxy mass is particularly pronounced for strong DLAs,  $N_{\text{HI}} > 10^{21} \text{ cm}^{-2}$ , the majority of which are associated with galaxies with  $M_{\star} \gtrsim 10^9 M_{\odot}$ . We analyzed different properties of galaxies that are linked to strong HI absorbers with different



HI column densities and found similar trends as we found for stellar mass: most strong HI absorbers are closely associated with galaxies that have relatively low halo masses, low SFRs and low HI masses, but strong DLAs (i.e.,  $N_{\text{HI}} > 10^{21} \text{ cm}^{-2}$ ) are typically linked to more massive galaxies with significantly higher halo masses, SFRs and HI masses.

By analyzing subsets of strong HI absorbers for which the associated galaxies have specific properties, we found that observationally confirmed DLA-galaxy pairs that have measured mass or SFR, have impact parameters that are in good agreement with our predictions. We stress, however, that the majority of DLAs are predicted to be more closely associated with galaxies that are at smaller impact parameters, but are too faint to be detected with current surveys. Hence, the masses and impact parameters of the observed galaxy counterparts of DLAs are both biased high. This is consistent with the large number of non-detections in observational campaigns that searched for galaxies close to DLAs (e.g., Foltz et al., 1986; Smith et al., 1989; Lowenthal et al., 1995; Bunker et al., 1999; Prochaska et al., 2002; Kulkarni et al., 2006; Rahmani et al., 2010; Bouché et al., 2012).

In order to facilitate the comparison between cosmological simulations and observations, we provide statistics on DLA-galaxy pairs for a few different SFR thresholds. However, a proper comparison requires observational studies aiming to find galaxies close to DLAs, to report their detection limit, the maximum allowed velocity separation, and either impact parameter of the nearest detectable galaxy or, in the case of non-detection (which must always be reported), the maximum impact parameter that has been searched. For the few studies that report such information (e.g., Teplitz et al., 1998; Mannucci et al., 1998), we found a good agreement with our simulation.

Interestingly, some recent observational studies indicate that strong HI absorbers at  $z \sim 2 - 3$  are associated with surprisingly massive galaxies. In particular, Rudie et al. (2012) studied the distribution of HI absorbers around a sample of rest-frame UV-selected Lyman-Break galaxies (LBG) with typical masses of  $M_{\star} > 10^{10} M_{\odot}$  at  $z \sim 2 - 3$  and found that nearly half of absorbers with  $10^{16} < N_{\text{HI}} < 10^{17} \text{ cm}^{-2}$  reside within a line-of-sight velocity difference of  $300 \text{ km s}^{-1}$  and a transverse separation of 300 proper kpc from a LBG, a region they labelled the circumgalactic medium. This result appears to contradict our finding that most strong HI absorbers are associated with galaxies with  $M_{\star} \lesssim 10^8 M_{\odot}$ . We demonstrated, however, that even though the absorbers are physically most closely associated with low-mass galaxies, these galaxies cluster strongly around galaxies as massive as LBGs, which is sufficient to reproduce the observations of Rudie et al. (2012). Moreover, we noted that the clustering does not even need to be that strong: Since 300 proper kpc and  $300 \text{ km s}^{-1}$  exceed the virial radius and the circular velocity of the haloes thought to host LBGs, by more than 300 and 50 per cent, respectively (e.g., Rakic et al., 2012, 2013), nearly all the volume of the “circumgalactic medium” lies beyond the virial radius if we employ the definition of Rudie et al. (2012).

Future deep observational surveys using new instruments (e.g., MUSE; Bacon et al., 2010) will improve this situation by detecting fainter galaxies. However, missing faint galaxies is a generic feature for any survey that has a finite detection limit and that takes an absorber-centered point of view. The incompleteness problem can be overcome by taking a galaxy-centered point of view, but this approach is inefficient for rare absorbers such as the interesting strong H<sub>I</sub> systems we study here. Moreover, while galaxy-centered surveys can measure the statistical distribution of absorbers (such as their covering factor), we still need to avoid interpreting the selected galaxy as the counterpart to any absorber that is detected. Absorber-centered surveys will probably remain the most efficient way to build up large numbers of galaxy-DLA pairs. Even with modest detection limits, such surveys provide highly valuable constraints on the relation between absorbers and galaxies, provided all non-detections are reported and that the detection limits and the maximum possible impact parameter are clearly specified.

## Acknowledgments

We would like to thank A. Pawlik, X. Prochaska, M. Raičević and M. Shirazi for useful discussion, reading an earlier version of this chapter and providing us with comments that improved the text. The simulations presented here were run on the Cosmology Machine at the Institute for Computational Cosmology in Durham (which is part of the DiRAC Facility jointly funded by STFC, the Large Facilities Capital Fund of BIS, and Durham University) as part of the Virgo Consortium research programme. This work was sponsored with financial support from the Netherlands Organization for Scientific Research (NWO), also through a VIDI grant and an NWO open competition grant. We also benefited from funding from NOVA, from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement 278594-GasAroundGalaxies and from the Marie Curie Training Network CosmoComp (PITN-GA-2009-238356).

## References

- Adelberger, K. L., Steidel, C. C., Shapley, A. E., & Pettini, M. 2003, *ApJ*, 584, 45  
Adelberger, K. L., Shapley, A. E., Steidel, C. C., et al. 2005, *ApJ*, 629, 636  
Aguirre, A., Dow-Hygelund, C., Schaye, J., & Theuns, T. 2008, *ApJ*, 689, 851  
Altay, G., Theuns, T., Schaye, J., Crighton, N. H. M., & Dalla Vecchia, C. 2011, *ApJL*, 737, L37  
Altay, G., Theuns, T., Schaye, J., Booth, C. M., & Dalla Vecchia, C. 2013, *arXiv:1307.6879*  
Bacon, R., Accardo, M., Adjali, L., et al. 2010, *Proceedings of the SPIE*, 7735  
Blitz, L., & Rosolowsky, E. 2006, *ApJ*, 650, 933



- Booth, C. M., & Schaye, J. 2009, *MNRAS*, 398, 53
- Bouché, N., Murphy, M. T., Péroux, C., et al. 2012, *MNRAS*, 419, 2
- Bunker, A. J., Warren, S. J., Clements, D. L., Williger, G. M., & Hewett, P. C. 1999, *MNRAS*, 309, 875
- Cen, R. 2012, *ApJ*, 748, 121
- Chabrier, G. 2003, *PASP*, 115, 763
- Christensen, L., Wisotzki, L., Roth, M. M., et al. 2007, *A&A*, 468, 587
- Dalla Vecchia, C., & Schaye, J. 2008, *MNRAS*, 387, 1431
- Di Matteo, T., Springel, V., & Hernquist, L. 2005, *Nature*, 433, 604
- Djorgovski, S. G., Pahre, M. A., Bechtold, J., & Elston, R. 1996, *Nature*, 382, 234
- Dolag, K., Borgani, S., Murante, G., & Springel, V. 2009, *MNRAS*, 399, 497
- Erkal, D., Gnedin, N. Y., & Kravtsov, A. V. 2012, *ApJ*, 761, 54
- Foltz, C. B., Chaffee, F. H., Jr., & Weymann, R. J. 1986, *AJ*, 92, 247
- Font-Ribera, A., Miralda-Escudé, J., Arnau, E., et al. 2012, *JCAP*, 11, 59
- Fumagalli, M., O'Meara, J. M., Prochaska, J. X., & Kanekar, N. 2010, *MNRAS*, 408, 362
- Fumagalli, M., Prochaska, J. X., Kasen, D., et al. 2011, *MNRAS*, 418, 1796
- Fynbo, J. U., Møller, P., & Warren, S. J. 1999, *MNRAS*, 305, 849
- Fynbo, J. P. U., Laursen, P., Ledoux, C., et al. 2010, *MNRAS*, 408, 2128
- Fynbo, J. P. U., Ledoux, C., Noterdaeme, P., et al. 2011, *MNRAS*, 413, 2481
- Fynbo, J. P. U., Geier, S., Christensen, L., et al. 2013, arXiv:1306.2940
- 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
- Haardt F., Madau P., 2001, in *Clusters of Galaxies and the High Redshift Universe Observed in X-rays*, Neumann D. M., Tran J. T. V., eds.
- Haas, M. R., Schaye, J., Booth, C. M., et al. 2012, arXiv:1211.1021
- Haehnelt, M. G., Steinmetz, M., & Rauch, M. 1998, *ApJ*, 495, 647
- Hennawi, J. F., & Prochaska, J. X. 2007, *ApJ*, 655, 735
- Krogager, J.-K., Fynbo, J. P. U., Møller, P., et al. 2012, *MNRAS*, 424, L1
- Krogager, J.-K., Fynbo, J. P. U., Ledoux, C., et al. 2013, arXiv:1304.4231
- Komatsu, E., et al. 2011, *ApJS*, 192, 18
- Krumholz, M. R., Ellison, S. L., Prochaska, J. X., & Tumlinson, J. 2009, *ApJL*, 701, L12
- Kulkarni, V. P., Woodgate, B. E., York, D. G., et al. 2006, *ApJ*, 636, 30
- Lara-López, M. A., Cepa, J., Bongiovanni, A., et al. 2010, *A&A*, 521, L53
- Lowenthal, J. D., Hogan, C. J., Green, R. F., et al. 1995, *ApJ*, 451, 484
- Mannucci, F., Thompson, D., Beckwith, S. V. W., & Williger, G. M. 1998, *ApJL*, 501, L11
- Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, *MNRAS*, 408, 2115
- McDonald, P., & Miralda-Escudé, J. 1999, *ApJ*, 519, 486
- Moller, 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

## REFERENCES

---

- Möller, P., Fynbo, J. P. U., & Fall, S. M. 2004, *A&A*, 422, L33
- Monier, E. M., Turnshek, D. A., & Rao, S. 2009, *MNRAS*, 397, 943
- Nagamine, K., Springel, V., & Hernquist, L. 2004, *MNRAS*, 348, 421
- Noterdaeme, P., Laursen, P., Petitjean, P., et al. 2012, *A&A*, 540, A63
- Pawlik, A. H., & Schaye, J. 2008, *MNRAS*, 389, 651
- Pawlik, A. H., & Schaye, J. 2011, *MNRAS*, 412, 1943
- Péroux, C., Bouché, N., Kulkarni, V. P., York, D. G., & Vladilo, G. 2011, *MNRAS*, 410, 2237
- Pontzen, A., Governato, F., Pettini, M., et al. 2008, *MNRAS*, 390, 1349
- Prochaska, J. X., & Wolfe, A. M. 1997, *ApJ*, 487, 73
- Prochaska, J. X., & Wolfe, A. M. 1998, *ApJ*, 507, 113
- Prochaska, J. X., Gawiser, E., Wolfe, A. M., et al. 2002, *AJ*, 123, 2206
- Prochaska, J. X., Gawiser, E., Wolfe, A. M., Cooke, J., & Gelino, D. 2003, *ApJS*, 147, 227
- Prochaska, J. X., & Wolfe, A. M. 2009, *ApJ*, 696, 1543
- Prochaska, J. X., Hennawi, J. F., & Simcoe, R. A. 2013, *ApJL*, 762, L19
- Rahmani, H., Srianand, R., Noterdaeme, P., & Petitjean, P. 2010, *MNRAS*, 409, L59
- Rahmati, A., Pawlik, A. H., Raičević, M., & Schaye, J. 2013a, *MNRAS*, 430, 2427
- Rahmati, A., Schaye, J., Pawlik, A. H., & Raičević, M. 2013b, *MNRAS*, 431, 2261
- Rakic, O., Schaye, J., Steidel, C. C., et al. 2013, arXiv:1306.1563
- Rakic, O., Schaye, J., Steidel, C. C., & Rudie, G. C. 2012, *ApJ*, 751, 94
- Rao, S. M., Belfort-Mihalyi, M., Turnshek, D. A., et al. 2011, *MNRAS*, 416, 1215
- Razoumov, A. O., Norman, M. L., Prochaska, J. X., & Wolfe, A. M. 2006, *ApJ*, 645, 55
- Rudie, G. C., Steidel, C. C., Trainor, R. F., et al. 2012, *ApJ*, 750, 67
- Schaye, J. 2001a, *ApJL*, 559, L1
- Schaye, J. 2001c, *ApJL*, 562, L95
- Schaye, J., & Dalla Vecchia, C. 2008, *MNRAS*, 383, 1210
- Schaye, J., Dalla Vecchia, C., Booth, C. M., et al. 2010, *MNRAS*, 402, 1536
- Smith, H. E., Cohen, R. D., Burns, J. E., Moore, D. J., & Uchida, B. A. 1989, *ApJ*, 347, 87
- Springel, V. 2005, *MNRAS*, 364, 1105
- Steidel, C. C., Pettini, M., & Hamilton, D. 1995, *AJ*, 110, 2519
- Steidel, C. C., Erb, D. K., Shapley, A. E., et al. 2010, *ApJ*, 717, 289
- Tescari, E., Viel, M., Tornatore, L., & Borgani, S. 2009, *MNRAS*, 397, 411
- Teplitz, H. I., Malkan, M., & McLean, I. S. 1998, *ApJ*, 506, 519
- van de Voort, F., Schaye, J., Altay, G., & Theuns, T. 2012a, *MNRAS*, 421, 2809
- van de Voort, F., & Schaye, J. 2012b, *MNRAS*, 423, 2991
- Wiersma, R. P. C., Schaye, J., Theuns, T., Dalla Vecchia, C., & Tornatore, L. 2009a, *MNRAS*, 399, 574
- Wiersma, R. P. C., Schaye, J., & Smith, B. D. 2009b, *MNRAS*, 393, 99
- Wolfe, A. M., Turnshek, D. A., Smith, H. E., & Cohen, R. D. 1986, *ApJS*, 61, 249
- Wolfe, A. M., Gawiser, E., & Prochaska, J. X. 2005, *ARA&A*, 43, 861

Wolfe, A. M., Prochaska, J. X., Jorgenson, R. A., & Rafelski, M. 2008, *ApJ*, 681, 881

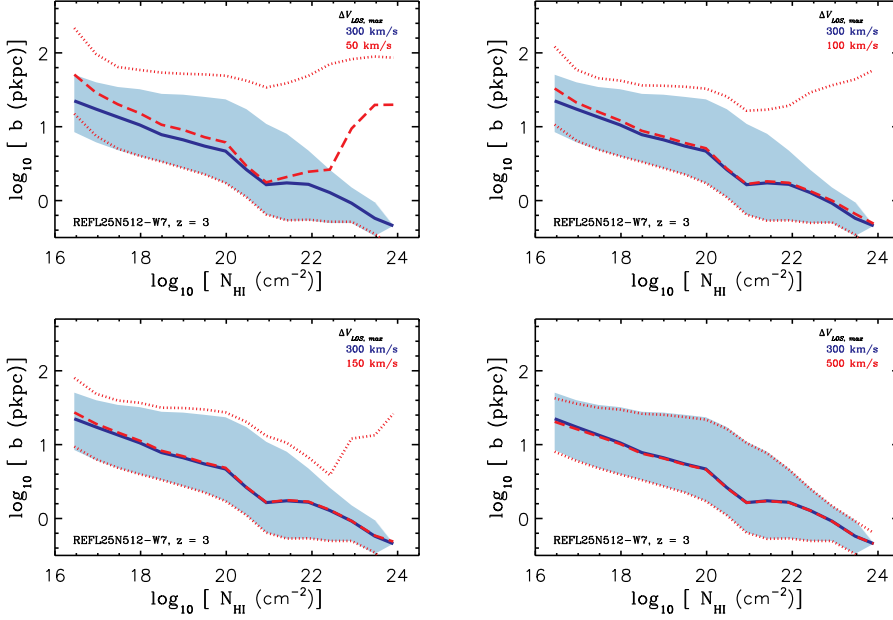
Yanny, B., York, D. G., & Williams, T. B. 1990, *ApJ*, 351, 377

## Appendix A: Choosing the maximum allowed LOS Velocity difference

As discussed in §4.2.4, for associating the H I absorbers to galaxies that are close to them we take into account their relative LOS velocities. This allows us to take out systems that appear to be close in projection but are apart by large physical distances. In analogy to observational studies, we calculate LOS velocities of H I absorbers and galaxies. If the difference between these two LOS velocities be larger than a minimum value, then we do not associate them as counterparts even though they are very close in projection. One might think that the above mentioned allowed minimum velocity difference between H I absorbers and galaxies, should be as small as possible to minimize the projection effect. In reality, however, the peculiar velocity of the H I gas around galaxies is not negligible and is controlled by complex processes like accretion and outflows. If the allowed LOS velocity differences are too small, peculiar velocity of H I absorbers around galaxies would be miss-interpreted as large LOS sight distances. As shown in Figure 4.9, our exercise show that the median impact parameter of absorbers as a function of  $N_{\text{HI}}$  is converged for maximum LOS velocity difference of  $\Delta V_{\text{LOS, max}} > 100$  km/s and the scatter around the median impact parameters is converged for  $\Delta V_{\text{LOS, max}} > 300$  km/s. Therefore, we adopt  $\Delta V_{\text{LOS, max}} = 300$  km/s which is also consistent with recent observational studies (e.g., Rakic et al., 2012; Rudie et al., 2012).

## Appendix B: Impact of feedback

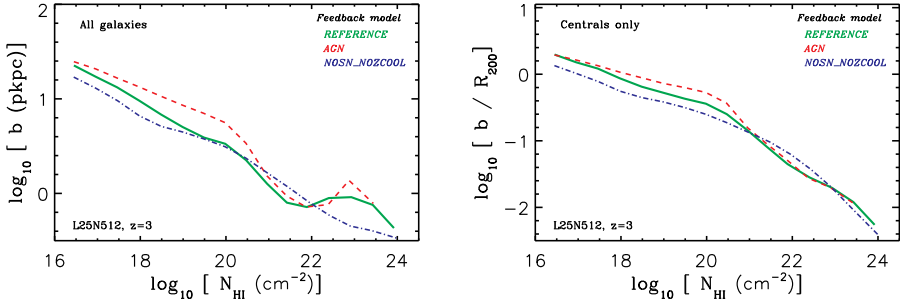
The evolution of gas and stars is determined by complex baryonic interactions that are often modeled in cosmological simulations by combining various physically motivated and empirical ingredients. In this context, different feedback mechanisms can change both the distribution of gas around galaxies with a given mass and the abundance of galaxies with different masses (e.g., van de Voort & Schaye, 2012b; Haas et al., 2012). As a result, the strength and details of various feedback mechanisms change the distribution of the H I absorbers (Altay et al., 2013) and may also affect the relative distribution of H I absorbers and galaxies. To quantify the impact of feedback on our results, we compare the relation we found in our reference model between the impact parameter of absorbers and their  $N_{\text{HI}}$  to the same relation in similar simulations that use different feedback prescriptions. Figure 4.10 shows this comparison between the



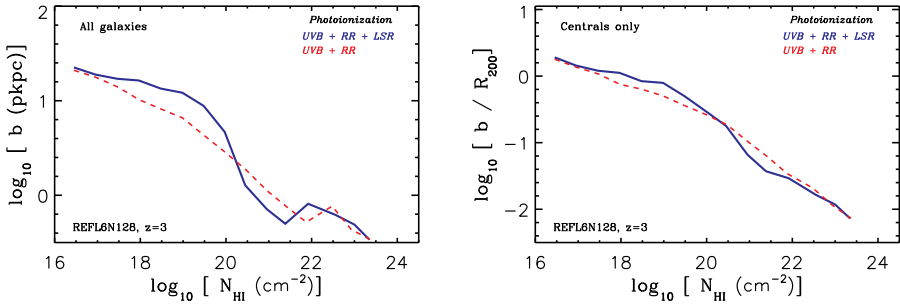
**Figure 4.9:** The impact of changing the maximum allowed LOS velocity difference between HI absorbers and galaxies to associate them with each other. The blue solid curve in all panels shows the median impact parameter of absorbers in our simulation at  $z = 3$  where our fiducial value of  $\Delta V_{\text{LOS, max}} = 300$  km/s is adopted and the shaded area around the blue solid curve shows the 15% – 85% percentiles. In each panel, the same result but using a different  $\Delta V_{\text{LOS, max}}$  is shown by red dashed and dotted curves which, respectively, show the median and the 15% – 85% percentiles. The  $\Delta V_{\text{LOS, max}}$  that is compared with our fiducial choice of 300 km/s is varying from 50 km/s from the top-left panel to 500 km/s in the bottom-right panel. While the median impact parameters are converged for  $\Delta V_{\text{LOS, max}} > 100$  km/s, both the median impact parameters and scatter around it are converged for  $\Delta V_{\text{LOS, max}} > 300$  km/s.

reference simulation at  $z = 3$  and two bracketing models that use AGN feedback and do not have SNe feedback and metal cooling. The solid green curve in the left panel of Figure 4.10 shows the reference model while the red dashed curve and blue dot-dashed curves respectively indicate the simulation with AGN and the simulation without SNe feedback and metal cooling (NOSN\_NOZCOOL). The relation between the normalized impact parameter and  $N_{\text{HI}}$  is shown in the right panel of Figure 4.10, where only central galaxies are taken into account for the matching process.

The absorbers with  $N_{\text{HI}} \lesssim 10^{21} \text{ cm}^{-2}$  in the AGN simulation have typical impact parameters which are larger than in the reference model. At very high HI column densities however, the two models are similar except that there are fewer very high  $N_{\text{HI}}$  systems in the AGN simulation. The same trend is seen



**Figure 4.10:** The (normalized) impact parameter as a function of  $N_{\text{HII}}$  at  $z = 3$  for simulations with different feedback models is shown in the (right) left panel. The green solid, red dashed and blue dot-dashed curves respectively show the reference simulation, the impact of adding AGN feedback and the result of turning off SNe feedback and metal cooling. While the impact parameters are sensitive to the adopted feedback prescription, the differences are much smaller than the intrinsic scatter caused by the complex geometry of gas distribution around galaxies with a wide range of sizes.



**Figure 4.11:** The (normalized) impact parameter as a function of  $N_{\text{HII}}$  at  $z = 3$  for the *REFL06N128* simulation with different photoionization models is shown in the (right) left panel. The red dashed curve shows the result when only the UVB and recombination radiations (RR) are present while the blue solid curve indicates the result of including local stellar radiation (LSR). The data points in the left panel are observations. Photoionization from local stellar radiation reduces the impact parameter of DLAs by up to 50% and increases the typical impact parameter of Lyman Limit systems by similar amount.

in the normalized impact parameters and is consistent with previous studies implying that AGN feedback pushes the gas away from the center of haloes by preventing the rapid cooling of the accreted gas and moderates the formation of the most massive galaxies which are the main contributors to very high HI column densities (e.g., Di Matteo et al., 2005; Booth & Schaye, 2009; Schaye et al., 2010; van de Voort & Schaye, 2012b; Haas et al., 2012). We also found that the contribution of galaxies with different stellar masses and SFRs as a function of  $N_{\text{HI}}$  in the AGN simulation is very similar to the reference model (not shown). The only difference between the two simulations in this context is that the number of strong HI absorbers in very massive galaxies (with  $M_{\star} \gtrsim 10^{10} M_{\odot}$ ;  $\text{SFR} \gtrsim 1 M_{\odot} \text{ yr}^{-1}$ ) becomes smaller while the contribution of less massive galaxies (with  $M_{\star} \lesssim 10^9 M_{\odot}$ ;  $\text{SFR} \lesssim 10^{-1} M_{\odot} \text{ yr}^{-1}$ ) remains intact.

In contrast, turning off the SNe feedback and metal cooling reduces the typical impact parameters of HI absorbers at almost all HI column densities. This can be understood by noting that the absence of metal cooling reduces the amount of cold gas available for accretion into galaxies, forcing the dense gas to be found closer to the center of galaxies. Moreover, the lack of SNe feedback allows the gas to fall more regularly along well defined filaments without being redistributed by SNe driven outflows. The lack of SNe feedback also allows the formation of dense cores in the center of lower mass galaxies which are interrupted easier by SNe explosions. As a result, the contribution of smaller galaxies (with  $M_{\star} \lesssim 10^9 M_{\odot}$ ;  $\text{SFR} < 1 M_{\odot} \text{ yr}^{-1}$ ) to the abundance of high  $N_{\text{HI}}$  systems increases. This results in smaller impact parameter at high HI column densities compared to the reference simulation, while leave their normalized impact parameters nearly identical.

All the above mentioned differences in the impact parameters of strong HI absorbers due to variations in feedback are much smaller than the intrinsic scatter in the expected impact parameter at a given  $N_{\text{HI}}$  which is caused by the complex geometry of gas distribution around galaxies with different sizes. The strong anti-correlation between the impact parameter of absorbers and their HI column densities is present and remains similar despite large variations in feedback mechanisms. Therefore, we conclude that feedback variation has a minor impact on our main results, although observations agree better with models that have stronger feedbacks.

## **Appendix C: Impact of local stellar radiation**

After the reionization of the Universe at  $z \sim 6$ , ionizing background radiation produced by the radiation coming from all sources in the Universe keeps hydrogen atoms mostly ionized. While on large scale, e.g., in the intergalactic medium, this radiation is uniform, it becomes highly non-uniform close to radiation sources. In particular, as was shown in Rahmati et al. (2013b), for absorbers with  $N_{\text{HI}} \gtrsim 10^{17} \text{ cm}^{-2}$  the radiation from local stellar radiation becomes import-

ant. Since we study the distribution of such strong H I absorbers, it is important to investigate the impact of local stellar radiation on the results of this work, where we neglect it.

The impact of local stellar radiation on the spatial distribution of strong H I absorbers for the *REFL06N128* simulation at  $z = 3$  is shown in Figure 4.11. The blue solid curve shows the result of radiative transfer calculation by taking into account the photoionization of the UVB, recombination radiation and local stellar radiation as explained in Rahmati et al. (2013b), while the red dashed curve indicates the result of including only the photoionization of the UVB and recombination radiation in the same simulation. As can be seen by comparing the two curves in the left panel of Figure 4.11, local stellar radiation can change the impact parameters of strong H I absorbers by up to 50%. The median impact parameters of DLAs with  $N_{\text{HI}} \sim 10^{21} \text{ cm}^{-2}$  is reduced because their H I column densities decreases due to additional photoionization by local stellar radiation. For Lyman Limit systems on the other hand, the ionization from local stellar radiation mainly affect systems that are closer to the galaxies compared to absorbers that are at larger distances. This results in an increase in the median impact parameter of absorbers at a given  $N_{\text{HI}}$  by decreasing the H I column density of absorbers at shorter impact parameters. At lower H I column densities (i.e.,  $N_{\text{HI}} \lesssim 10^{17} \text{ cm}^{-2}$ ) where the effect of local stellar radiation is negligible (Rahmati et al., 2013b), the impact parameters remain intact<sup>6</sup>. We conclude that local stellar radiation is not changing the results we present in this work.

## Appendix D: Resolution tests

We use three different cosmological simulations that have identical box sizes, but different resolutions. These simulations are part of the OWLS project (Schaye et al., 2010) and have cosmological parameters that are consistent with WMAP year-3 values (i.e.,  $\{\Omega_{\text{m}} = 0.238, \Omega_{\text{b}} = 0.0418, \Omega_{\Lambda} = 0.762, \sigma_8 = 0.74, n_{\text{s}} = 0.951, h = 0.73\}$ ), slightly different from our fiducial cosmology in this work. The simulation with the highest resolution (*REFL25N512*) has identical mass resolution and box size compared to the simulation we use in this work and the other two simulations have 8 times (*REFL25N256*) and 64 time (*REFL25N128*) lower resolutions (see Table 4.2 for more details).

As we showed in Rahmati et al. (2013a), the H I column density distribution function is converged for LLS and most DLAs at the resolution that we use in this work. The position of galaxies in cosmological simulations is determined by the distribution of overdensities and therefore, is not expected to be highly sensitive to the resolution. This is not true for the number of galaxies that are

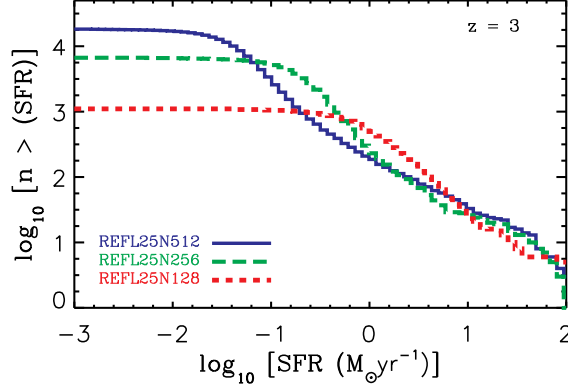
<sup>6</sup>We note that using a small simulation box results in underproducing the strong H I absorbers due to missing very massive galaxies (Rahmati et al., 2013a). Given that the contribution of very massive galaxies to the total abundance of absorbers increases with increasing the  $N_{\text{HI}}$  (see Figure 4.4 and 4.5), missing them in the small simulation box allows the smaller galaxies to be the main DLA counterparts and hence decreases the typical impact parameters of strong H I absorbers.



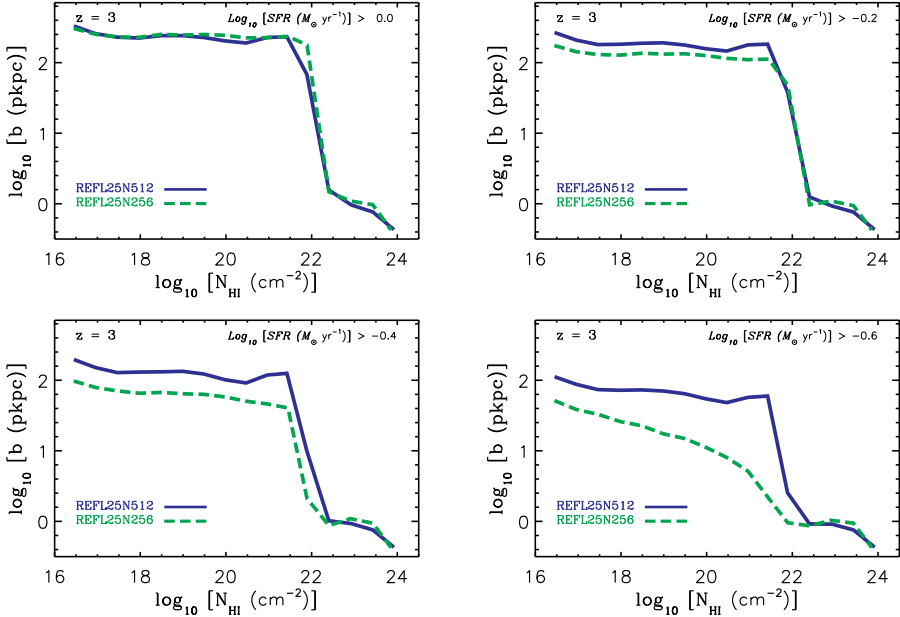
**Table 4.2:** List of cosmological simulations used in this work. The detailed description of the model ingredients are discussed in Schaye et al. (2010). From left to right the columns show: simulation identifier; comoving box size; number of dark matter particles (there are equally many baryonic particles); initial baryonic particle mass; dark matter particle mass; comoving (Plummer-equivalent) gravitational softening; maximum physical softening; final redshift; remarks about the used model, cosmology and the use of explicit radiative transfer calculations instead of fitting function for the HI calculations (RT).

Simulation	$L$ ( $h^{-1}\text{Mpc}$ )	$N$	$m_b$ ( $h^{-1}M_\odot$ )	$m_{\text{dm}}$ ( $h^{-1}M_\odot$ )	$\epsilon_{\text{com}}$ ( $h^{-1}\text{kpc}$ )	$\epsilon_{\text{prop}}$ ( $h^{-1}\text{kpc}$ )	$z_{\text{end}}$	Model
<i>REFL06N128</i>	6.00	$128^3$	$1.4 \times 10^6$	$6.3 \times 10^6$	1.95	0.50	0	REF, WMAP7 cosmology, RT
<i>REFL25N512-W7</i>	25.00	$512^3$	$1.4 \times 10^6$	$6.3 \times 10^6$	1.95	0.50	2	REF, WMAP7 cosmology
<i>REFL25N512</i>	25.00	$512^3$	$1.4 \times 10^6$	$6.3 \times 10^6$	1.95	0.50	1	REF
<i>AGN</i>	25.00	$512^3$	$1.4 \times 10^6$	$6.3 \times 10^6$	1.95	0.50	2	with AGN
<i>NOSN_NOZCOOL</i>	25.00	$512^3$	$1.4 \times 10^6$	$6.3 \times 10^6$	1.95	0.50	2	w/o SN, w/o metal cooling
<i>REFL25N256</i>	25.00	$256^3$	$1.1 \times 10^7$	$5.1 \times 10^7$	3.91	1.00	2	REF
<i>REFL25N128</i>	25.00	$128^3$	$8.7 \times 10^7$	$4.1 \times 10^8$	7.81	2.00	0	REF

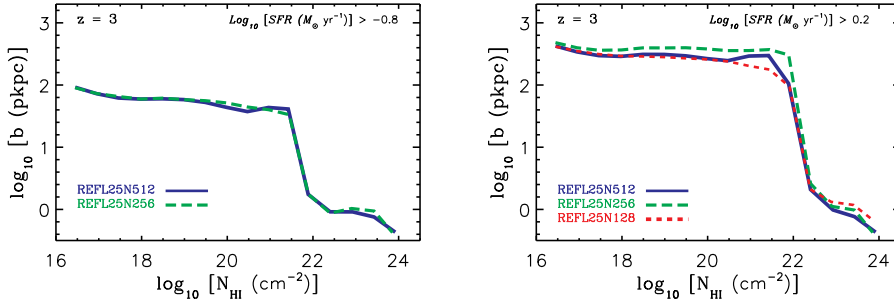




**Figure 4.12:** Cumulative number of galaxies that are resolved in simulations at  $z = 3$  with different resolutions as a function of their SFR. Blue solid, green long-dashed and red dashed curves show the *REFL25N512*, *REFL25N256* and *REFL25N128* simulations, respectively.



**Figure 4.13:** The resolution dependence of  $b - N_{\text{HI}}$  relation at different SFR thresholds. The blue solid and green dashed curves represent the *REFL25N512* and the *REFL25N256* simulations, respectively. Panels show the impact parameter of absorbers as a function  $N_{\text{HI}}$  by including only galaxies that have SFR above a certain threshold. The SFR thresholds from top-left to bottom-right are 1, 0.63, 0.4 and 0.2  $M_{\odot} \text{ yr}^{-1}$ , respectively. The results are similar in the two simulations for  $\text{SFR} > 0.4 M_{\odot} \text{ yr}^{-1}$ .



**Figure 4.14:** Impact parameter of HI absorbers as a function of their  $N_{\text{HI}}$  for different SFR thresholds. In the left panel, the blue solid curve shows the result for the *REFL25N512* simulation where only galaxies with  $\text{SFR} > 0.16 M_{\odot} \text{ yr}^{-1}$  are taken into account. The green long-dashed curve shows the result for the *REFL25N256* simulation where the SFR threshold is such that the cumulative number density of galaxies is matched to that of galaxies with  $\text{SFR} > 0.16 M_{\odot} \text{ yr}^{-1}$  in the *REFL25N512* simulation. In the right panel, the blue solid curve shows the  $b - N_{\text{HI}}$  relation for the *REFL25N512* simulation where only galaxies with  $\text{SFR} > 1.6 M_{\odot} \text{ yr}^{-1}$  are taken into account. The green long-dashed and red dashed curves respectively show the result in the *REFL25N256* and *REFL25N128* simulations where the SFR thresholds are chosen such that the total number density of galaxies that are taken into account is matched to that of galaxies with  $\text{SFR} > 1.6 M_{\odot} \text{ yr}^{-1}$  in the *REFL25N512* simulation. This shows that for a fixed total number density of galaxies the relation between impact parameters and the HI column density of absorbers is independent of the resolution.

resolved in simulations and as the resolution increases, the number of structures that are resolved in a simulation also increases. This is shown in Figure 4.12 which shows the cumulative distribution of number of objects that are identified as bound structures in our simulations<sup>7</sup>. Comparing the blue solid and green long-dashed curves in this figure shows that the number of galaxies that have  $\text{SFR} > 1 M_{\odot} \text{ yr}^{-1}$  is identical in the two simulations with highest resolutions. Together with the converged HI column density distribution function, this result implies that the relation between the impact parameter of HI absorbers and galaxies with  $\text{SFR} \gtrsim 1 M_{\odot} \text{ yr}^{-1}$  is also converged in the *REFL25N256* simulation. Indeed, as Figure 4.13 shows, this is the case for  $\text{SFR} > 0.4 M_{\odot} \text{ yr}^{-1}$ . This suggests that the  $b - N_{\text{HI}}$  relation is converged in the *REFL25N512* simulation for  $\text{SFR} > 6 \times 10^{-2} M_{\odot} \text{ yr}^{-1}$ . By increasing the number density of galaxies in a simulation, one would expect a decrease in the average distance between HI absorbers and galaxies. This suggests that the distribution of HI absorbers relative to galaxies to be primarily sensitive to the cumulative number density of

<sup>7</sup>We note that contrary to the *REFL25N512* simulation in which cosmological parameters are based on WMAP year-3 values, our reference simulation in this work is based on WMAP year-7 cosmological parameters. As a result its cumulative distribution of galaxies flattens at lower SFR compared to what is shown by the blue solid curve in Figure 4.12.

all galaxies. As a result, one might expect to retrieve the same relation between impact parameters and  $N_{\text{HI}}$  of absorbers if the total number density of galaxies be the same in simulations with different resolution. As Figure 4.14 shows, this is indeed true in our simulations where in each panel we choose different SFR thresholds depending on the resolution to match the total number(density) of galaxies above the SFR threshold for all the simulations. This result suggests that if one keeps the total number density of galaxies we use in our study, increasing the resolution is not expected to change the  $b - N_{\text{HI}}$  relation and other conclusions we derived in this work. In this work we used the relatively low SFR threshold of  $\text{SFR} > 4 \times 10^{-3} M_{\odot} \text{ yr}^{-1}$  to include as many bound (sub) structure as possible in our analysis. The total number density of galaxies that are selected in our reference simulation (i.e., *REFL25N512-W7*) by this criterium is 0.5 galaxy per comoving  $\text{Mpc}^3$  (i.e., equivalent to 31.5 galaxy per proper  $\text{Mpc}^3$ ). We note, however, that the SFR threshold that correspond to this total number density at higher resolutions is expected to be lower than  $4 \times 10^{-3} M_{\odot} \text{ yr}^{-1}$ . This can be seen in Figure 4.12 which shows that at a fixed cumulative number density, the SFR of galaxies in the *REFL25N256* simulation that have  $\text{SFR} < 1 M_{\odot} \text{ yr}^{-1}$  decreases by increasing the resolution.



# 5

## GENESIS OF THE DUSTY UNIVERSE: MODELING SUBMILLIMETRE SOURCE COUNTS

We model the evolution of infrared galaxies using a phenomenological approach to match the observed source counts at different infrared wavelengths. In order to do that, we introduce a new algorithm for reproducing source counts which is based on direct integration of probability distributions rather than using Monte-Carlo sampling. We also construct a simple model for the evolution of the luminosity function and the colour distribution of infrared galaxies which utilizes a minimum number of free parameters; we analyze how each of these parameters is constrained by observational data. The model is based on pure luminosity evolution, adopts the Dale & Helou Spectral Energy Distribution (SED) templates, allowing for evolution in the infrared luminosity-colour relation, but does not consider Active Galactic Nuclei as a separately evolving population. We find that the  $850\ \mu\text{m}$  source counts and redshift distribution depend strongly on the shape of the luminosity evolution function, but only weakly on the details of the SEDs. Based on this observation, we derive the best-fit evolutionary model using the  $850\ \mu\text{m}$  counts and redshift distribution as constraints. Because of the strong negative  $K$ -correction combined with the strong luminosity evolution, the fit has considerable sensitivity to the sub- $L_*$  population at high redshift, and our best-fit shows a flattening of the faint end of the luminosity function towards high redshifts. Furthermore, our best-fit model requires a colour evolution which implies the typical dust temperatures of objects with the same luminosities to decrease with redshift. We then compare our best-fit model to observed source counts at shorter and longer wavelengths which indicates our model reproduces the  $70\ \mu\text{m}$  and  $1100\ \mu\text{m}$  source counts remarkably well, but under-produces the counts at intermediate wavelengths. Analysis reveals that the discrepancy arises at low redshifts, indicating that revision of the adopted SED library towards lower dust temperatures (at a fixed infrared luminosity) is required. This modification is equivalent to a population of cold galaxies existing at low redshifts, as also indicated by recent Herschel results, which are underrepresented in IRAS

sample. We show that the modified model successfully reproduces the source counts in a wide range of IR and submm wavelengths.

Alireza Rahmati, Paul van der Werf  
*Monthly Notices of the Royal Astronomical Society*  
Volume 418, Issue 1, pp. 176-194 (2011)

## 5.1 Introduction

Although dust is an unimportant component in the mass budget of galaxies, its presence radically alters the emergent spectrum of star forming galaxies. Since stars are born in dusty clouds, most of the energy of young stars is absorbed by dust particles, which are heated by the absorption process and radiate their energy away by thermal emission at infrared (IR) and submillimetre (submm) wavelengths. As a result, the infrared radiation of star forming galaxies is a useful measure of the massive star formation rate. Mainly due to the atmospheric opacity, the thermal radiation from dust could not be systematically studied until the launch of IRAS in the mid-1980s, which provided the first comprehensive view of the nearby dusty Universe. Ten years later COBE discovered the Cosmic Infrared Background (CIB) (Puget et al., 1996; Fixsen et al., 1998) and it turned out the observed power of the CIB is comparable to what can be deduced from the optical Universe. This was in contrast to the observations of local galaxies which suggest only one third of the energy output of galaxies is in IR bands (Lagache et al., 2005). Moreover, the relatively flat slope of the CIB at long wavelengths indicated a population of dusty galaxies which are distributed over a wide range of redshifts (Gispert et al., 2000). Thanks to various large surveys performed with different satellites and ground based observatories, this background radiation is now partly resolved into point sources at different wavelengths. While at shorter wavelengths the emission is mainly coming from local and low redshift galaxies, longer wavelengths contain information about larger distances.

The shape of emission spectrum of warm dust particles resembles a modified blackbody spectrum with a peak varying with the typical dust temperature which is observed to be around  $T \sim 30 - 40\text{K}$  in galaxies; therefore the far-IR (FIR) and submm spectrum of a typical dusty galaxy consists of a peak at rest-frame wavelengths around  $\lambda \sim 100 - 200\mu\text{m}$  which drops on both sides (see Figure 5.1). While the presence of other emitters like polycyclic aromatic hydrocarbons (PAHs) and Active Galactic Nuclei (AGNs) complicates the shape of spectra at shorter wavelengths, at submm wavelengths the spectra behave like modified blackbodies and their amplitudes drop steeply. In fact, because of this steep falloff in the Rayleigh-Jeans tail of the Spectral Energy Distribution (SED), the observed flux density at a fixed submm wavelength can rise by moving the SED in redshift space (the so called *K*-correction), which counteracts the cosmological dimming. It turns out that the interplay between these two processes enables us to observe galaxies at submm wavelengths out to very high redshifts (see also the discussion in §5.6 and Figure 5.7).

After the first observations of SCUBA at  $850\mu\text{m}$  confirmed the importance of submm galaxies at high redshifts (Smail et al., 1997), there were many subsequent surveys using different instruments which explored different cosmological fields (Smail et al., 2002; Webb et al., 2003; Borys et al., 2003; Greve et al., 2004; Coppin et al., 2006; Bertoldi et al., 2007; Austermann et al.,

2010; Scott et al., 2010; Vieira et al., 2010) and also some surveys which used gravitational lensing techniques to extend the observable submm Universe to sub-mJy fluxes (Smail et al., 1997; Chapman et al., 2002; Knudsen et al., 2008; Johansson et al., 2011).

While low angular resolution makes individual identifications and spectroscopy a daunting task, the surface density of sources as a function of brightness (i.e., the source counts) can be readily analysed and contains significant information about the population properties and their evolution. One can assume simple smooth parametrized models for the evolution of dusty galaxies and relate their low redshift observed properties (e.g., the total IR luminosity function, which we simply call the luminosity function, or in short LF hereafter) to their source counts (Blain & Longair, 1993; Guiderdoni et al., 1997; Blain et al., 1999; Chary & Elbaz, 2001; Rowan-Robinson, 2001; Dole et al., 2003; Lagache et al., 2004; Lewis et al., 2005; Le Borgne et al., 2009; Valiante et al., 2009; Bethermin et al., 2011). Such backward evolution models usually combine SED templates and the low redshift properties of IR galaxies which are allowed to change with redshift based on an assumed parametric form, together with Monte-Carlo techniques to reproduce the observed source counts. These models are therefore purely phenomenological, and suppress the underlying physics. Their power lies exclusively in providing a parametrized description of statistical properties of the galaxy population under study, and the evolution of these properties. The results of such modeling provides a description of the constraints that must be satisfied (in a statistical sense) and could be explored further by more physically motivated models such as hydrodynamical simulations embedded in a  $\Lambda$ CDM cosmology, with subgrid prescriptions for star formation and feedback (e.g. Schaye et al. (2010)). It is important to realize that backward evolution models are constrained only over limited observational parameter space (e.g., at particular observing wavelengths, flux levels, redshift intervals, etc.), and are ignorant of the laws of physics. Therefore it is dangerous to use them outside of their established validity ranges; in other words, these models have descriptive power but little explanatory power. As a result, one of the most instructive uses of these models is in analyzing where they fail to reproduce the observations, and how they can be modified to correct for these failures which implies that the underlying assumptions must be revised. This is the approach that we use in the present chapter.

Since nowadays many and various observational constraints exist (e.g. source counts at various IR and submm wavelengths, colour distributions, redshift distribution), backward evolution models can reach considerable levels of sophistication. However, early backward evolution models used only a few SED templates (or sometimes even only one SED) to represent the whole population of dusty galaxies. This approach neglects the fact that dusty galaxies are not identical and span a variety of dust temperatures and hence SED shapes. As a result, such models cannot probe the possible evolution in the SED properties of submm galaxies, for which increasing observational support has been found



during the last few years (Chapman et al., 2005; Pope et al., 2006; Chapin et al., 2009; Symeonidis et al., 2009; Seymour et al., 2010; Hwang et al., 2010).

In addition, existing backward evolution models typically use only “luminosity” and/or “density” evolution which respectively evolves the characteristic luminosity (i.e.,  $L_*$ ) and the amplitude of the LF without changing its shape. In other words, they do not consider any evolution in the shape of the luminosity function. Finally, the intrinsically slow Monte-Carlo methods used in many of the models make the search of parameter space for the best-fit model laborious and tedious.

In this chapter, we use a new and fast algorithm which is different from conventional Monte-Carlo based algorithms, for calculating the source counts for a given backward evolution model. We also use a complete library of IR SED templates which form a representative sample of observed galaxies, at least at low redshifts. This will enable us to address questions about the evolution of colour distribution of dusty galaxies during the history of the Universe. We also consider a new evolution form for the LF which allows us to constrain evolution in the shape of the LF in addition to quantifying its amplitude changes.

The structure of the chapter is as follows: in §5.2 we present different ingredients of our parametric colour-luminosity function (CLF) evolution model and introduce a new algorithm for the fast calculation of the source counts for a given CLF. Then, after choosing our observational constraints for 850  $\mu\text{m}$  objects in §5.3 we try to find a model consistent with those constraints by studying the sensitivity of the produced source counts to different parameters in §5.4. After finding a 850  $\mu\text{m}$ -constrained best-fit model, we test its performance in producing source counts at other wavelengths in §5.6 which leads us to a model capable of producing source counts in a wide range of IR and submm wavelengths. Finally, after discussing the astrophysical implications of our best-fit model in §5.7, we end the chapter with concluding remarks. Throughout this chapter, we use the standard  $\Lambda\text{CDM}$  cosmology with the parameters  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$  and  $h = 0.75$ .

## 5.2 Model ingredients

Although luminosity is the main parameter to differentiate between different galaxies, it is an integrated property of SED and cannot distinguish between different SED shapes corresponding to different physical processes taking place in different objects with similar outgoing integrated energy fluxes. Using colour indicators can resolve this degeneracy. Dale et al. (2001) demonstrated that the  $R(60, 100)$  which is defined as  $\log(S_{60\mu\text{m}}/S_{100\mu\text{m}})$ , is the best single parameter characterization of IRAS galaxies. Moreover, it was shown that IRAS galaxies exhibit a slowly varying correlation between  $R(60, 100)$  and luminosity such that objects with larger characteristic luminosities have warmer characteristic colours (Dale et al., 2001; Chapman et al., 2005; Chapin et al., 2009). Based on these facts,

we construct a model for the Colour-Luminosity Function (hereafter CLF). This model consists of the observed CLF in the local Universe and a parametric evolution function. Then, by adopting a suitable set of SED models, we calculate the source count of IR objects at different wavelengths and compare the results with observations.

### 5.2.1 CLF at $z = 0$

The local volume density of IRAS galaxies,  $\Phi_0(L, C)$ , can be parametrized as a function of total IR luminosity,  $L$ , and  $R(60, 100)$  colour,  $C$ , where the total infrared luminosity is calculated by integrating over the SED from 3 to 1100  $\mu\text{m}$  (Dale et al., 2001). Furthermore, it is possible to express the local colour-luminosity distribution as the product of the local luminosity function,  $\Phi_0(L)$ , and the local conditional probability of a galaxy having the colour  $C$  given the luminosity  $L$ ,  $P_0(C|L)$ ,

$$\Phi_0(L, C) = \Phi_0(L)P_0(C|L). \quad (5.1)$$

Since the IRAS galaxies represent an almost complete sample of IR galaxies out to the redshift  $\sim 0.1$  (Saunders et al., 1990), we use the luminosity and colour functions found by analyzing a flux limited sample of  $S_{60\mu\text{m}} > 1.2\text{Jy}$  IRAS galaxies (Chapman et al., 2003; Chapin et al., 2009). Chapman et al. (2003) and Chapin et al. (2009) analyze this sample which covers most of the sky, using an accessible volume technique for finding the LF and fit a dual power law function to the observed luminosity distribution. The parametric form of luminosity function based on Chapin et al. (2009) is given by

$$\Phi_0(L) = \rho_* \left(\frac{L}{L_*}\right)^{1-\alpha} \left(1 + \frac{L}{L_*}\right)^{-\beta}, \quad (5.2)$$

where  $L_* = 5.14 \times 10^{10} L_\odot$  is the characteristic knee luminosity,  $\rho_* = 1.22 \times 10^{-14} \text{Mpc}^{-3} L_\odot^{-1}$  is the number density normalization of the function at  $L_*$ , and  $\alpha = 2.59$  and  $\beta = 2.65$  characterize the power-laws at the faint ( $L < L_*$ ) and bright ( $L > L_*$ ) ends, respectively.

Chapman et al. (2003) and Chapin et al. (2009) also found a Gaussian representation for the colour distribution of IRAS galaxies

$$P_0(C | L) = \frac{1}{\sqrt{2\pi}\sigma_c} \exp\left[-\frac{1}{2} \times \left(\frac{C - C_0}{\sigma_c}\right)^2\right], \quad (5.3)$$

where  $C_0$  can be represented by a dual power law function

$$C_0 = C_* - \delta \log\left(1 + \frac{L'}{L}\right) + \gamma \log\left(1 + \frac{L}{L'}\right), \quad (5.4)$$

with  $C_* = -0.48$ ,  $\delta = -0.06$ ,  $\gamma = 0.21$  and  $L' = 3.2 \times 10^9 L_\odot$ , and the distribution width,  $\sigma_c$ , is expressed as

$$\sigma_c = \sigma_f (1 - 2^{-L'/L}) + \sigma_b (1 - 2^{-L/L'}), \quad (5.5)$$

where  $\sigma_f = 0.2$  and  $\sigma_b = 0.128$  (Chapin et al., 2009).

## 5.2.2 CLF evolution

The CLF introduced in the previous section, contains information about the distribution of IR galaxies only in the nearby Universe. For modeling the IR Universe at higher redshifts, its evolution must be modeled. The necessity of the CLF evolution with redshift is shown directly by the fact that the observed power of the the CIB is comparable to what can be deduced from the optical cosmic background. This cannot be explained by our understanding of the local Universe which indicates the infrared output of galaxies is only one third of their optical output (Lagache et al., 2005). Furthermore, several studies have shown for a fixed total IR luminosity the typical temperature of infrared sources is lower at higher redshifts (Chapman et al., 2005; Pope et al., 2006; Chapin et al., 2009; Symeonidis et al., 2009; Seymour et al., 2010; Hwang et al., 2010; Amblard et al., 2010), which advocates an IR colour evolution. Therefore, we need to adopt a reasonable form of evolution in the luminosity and colour distributions to reproduce correctly the CLF evolution of IR galaxies with redshift.

There are three different ways to evolve the local luminosity function of IR galaxies: (i) changing  $\rho_*$  with redshift (i.e., density evolution); (ii) changing  $L_*$  with redshift (i.e., luminosity evolution) and (iii) changing the bright and faint end slopes (i.e.,  $\alpha$  and  $\beta$ ) with redshift. While the density and luminosity evolution change the abundances of all sources independent of their luminosities, they leave the shape of the LF unchanged. Such models are called *translational* models since they amount to only a translation of the LF in parameter space. Going beyond translational models, a variation of the shape of LF with redshift (i.e., varying  $\alpha$  and  $\beta$  slopes in equation 5.2) can be used to change the relative contributions of bright and faint sources at different redshifts.

As discussed by Blain et al. (1999), luminosity and density evolution affect the CIB predicted by the models similarly, but luminosity evolution has a much stronger effect on the source counts. Combining source counts and integrated background can therefore distinguish between luminosity and density evolution. The result of this analysis shows luminosity evolution must strongly dominate over density evolution, since pure density evolution consistent with the observed 850  $\mu\text{m}$  source counts would overpredict the integrated background by a factor 50 to 100 (Blain et al., 1999). We therefore assume negligible density evolution. However, as we will show in §5.4, luminosity evolution is not sufficient to reproduce the correct source count and redshift distribution of submm galaxies, which forces us to drop the assumption of a purely translational LF evolution model. Moreover, there is no reason to assume that the slopes of LF at faint and bright ends remain the same at all redshifts. Therefore, we allow them to change in our model and introduce a redshift dependent LF which can be written as

$$\Phi(L, z) = \rho_* \left( \frac{L}{g(z)L_*} \right)^{1-\alpha_z} \left( 1 + \frac{L}{g(z)L_*} \right)^{-\beta_z}, \quad (5.6)$$

which is similar to equation (5.2) but now  $\alpha_z$  and  $\beta_z$  are changing with redshift

and  $L_*$  is multiplied by  $g(z)$ , which we refer to as the luminosity evolution function.

Since the average total-IR-luminosity density and its associated star formation density closely follow the luminosity evolution function, we choose a form of luminosity evolution which is similar to the observed evolution of the cosmic star formation history (see Hopkins & Beacon (2006)); in other words, we assume that the luminosity evolution is a function which is increasing at low redshifts and after reaching its maximum turns into a decreasing function of look-back time at high redshifts:

$$g(z) = \begin{cases} (1+z)^n & \text{if } z \leq z_a \\ (1+z_a)^n & \text{if } z_a < z \leq z_b \\ (1+z_a)^n(1+z-z_b)^m & \text{if } z_b < z \end{cases} \quad (5.7)$$

where  $n$ ,  $m$ ,  $z_a$  and  $z_b$  are constants. As we will show in §5.4.2, the source count is not strongly sensitive to the model properties at high redshifts. In other words, the sensitivity of the source count calculation to the difference between  $z_a$  and  $z_b$  is much less than its high sensitivity to the value of  $z_a$  itself. Moreover, the model outputs are not highly sensitive to the slope of  $g(z)$  at high redshifts if it remains negative (i.e.,  $m < 0$ ). Therefore, we can simplify the model by assigning suitable fixed values to  $z_b - z_a$  and  $m$ , without any significant change in its flexibility. We choose  $z_b - z_a = 1$  and  $m = -1$ , knowing that any different choice for those values can be compensated by a very small change in  $z_a$  or/and  $n$ . we will discuss this issue in more detail in §5.4.2.

We also adopt linear forms for changing  $\alpha_z$  and  $\beta_z$  with redshift

$$\alpha_z = 2.59 + a_\alpha z, \quad (5.8)$$

$$\beta_z = 2.65 + a_\beta z, \quad (5.9)$$

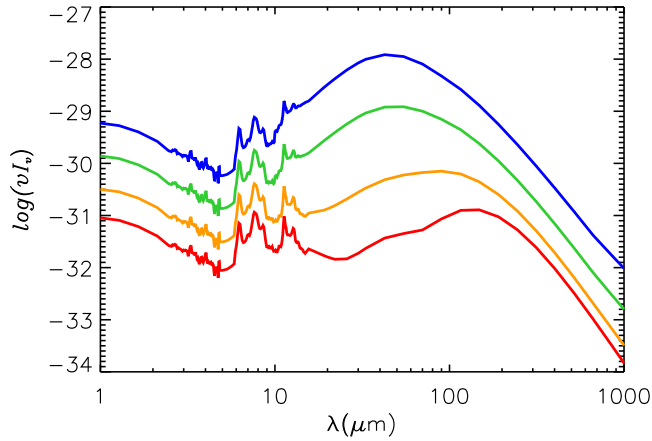
where  $a_\alpha$  and  $a_\beta$  are constants. We apply the slope evolution only for  $z < \frac{z_a+z_b}{2}$  noting that the observed source count is not very sensitive to the exact properties of our model at very high redshifts and extending the evolution of LF slopes to even higher redshifts does not change our results.

As mentioned earlier, some studies have shown that high redshift IR galaxies have lower dust temperatures than low redshift galaxies, at a fixed total IR luminosity. In our model, in order to evolve the colour distribution accordingly, we adopt a relation similar to Valiante et al. (2009) to shift the centre of the colour distribution for a given luminosity towards colours associated with lower dust temperatures (i.e., smaller  $R(60, 100)$  values) at higher redshifts (see Figure 5.12). In this way, the colour distribution of IR galaxies can be written as

$$P(C | L) = \frac{1}{\sqrt{2\pi}\sigma_c} \exp\left[-\frac{1}{2} \times \left(\frac{C - C'_0}{\sigma_c}\right)^2\right], \quad (5.10)$$

where  $\sigma_c$  is given by equation (5.5) and  $C'_0$  is

$$C'_0 = C_* - \delta \log\left(1 + \frac{L'(1+z)^w}{L}\right) + \gamma \log\left(1 + \frac{L}{L'(1+z)^w}\right), \quad (5.11)$$



**Figure 5.1:** Some SED samples from the template library of Dale & Helou (2002) which we use in our model. The blue, green, orange and red lines (from top to bottom) are respectively associated with colours  $C = 0.16, 0.04, -0.27$  and  $-0.54$ . The normalization is arbitrary.

where  $w$  is a constant. Although for colour evolution we adopt a similar formalism to Valiante et al. (2009), unlike them we allow  $w$  to vary as a free parameter which enables us to constrain the colour evolution as well.

Based on the above formalism, the general CLF of IR galaxies can be written as

$$\Phi(L, C, z) = \Phi(L, z)P(C|L), \quad (5.12)$$

where  $\Phi(L, z)$  and  $P(C|L)$  are defined by equations (5.6)-(5.11).

### 5.2.3 SED Model

Several studies have shown that AGNs do not dominate the FIR energy output of the Universe (Swinbank et al., 2004; Alexander et al., 2005; Chapman et al., 2005; Lutz et al., 2005; Valiante et al., 2007; Pope et al., 2008; Menendez-Delmestre et al., 2009; Fadda et al., 2010; Jauzac et al., 2010). Therefore, in order to avoid unnecessary complexity in our model we adopt a single population of galaxies, modeled with only one family of SEDs representing star forming galaxies.

This choice is justified for long wavelengths, for instance at  $850 \mu\text{m}$ , by considering the fact that at those wavelengths the AGN contribution to the observed flux is negligible. In other words, to be able to change the  $850 \mu\text{m}$  fluxes and source counts, AGN should be the dominant contributor to the SED at rest-frame wavelengths longer than  $\sim 200 \mu\text{m}$  which is highly unlikely.

On the other hand, excluding AGN contribution could be potentially important at shorter wavelengths: if we adopt a simple assumption where the AGN continuum is well represented by a simple torus model (Efstathiou et al., 1995; Valiante et al., 2009), then a strong enough AGN could create a bump, on top of the starburst SED, at rest-frame wavelength ranges  $\sim 10 - 50 \mu\text{m}$  (see Figure 2 in Efstathiou et al. (1995) and Figure 9 in Valiante et al. (2009)). This feature, together with increasing AGN luminosity with redshift could modify the observed source counts at observed wavelengths shorter than  $200 \mu\text{m}$  but still is unlikely to affect submm counts.

However, Mullaney et al. (2011) recently used the deepest available Herschel survey and showed that for a sample of X-ray selected AGNs up to  $z \sim 3$ , the observed  $100 \mu\text{m}$  and  $160 \mu\text{m}$  fluxes are not contaminated by AGN. Consequently, even if all the galaxies which contribute to the FIR and submm source counts host AGN, their observed fluxes is driven by star formation activity at wavelengths around  $100 \mu\text{m}$  or longer.

In order to get accurate SED templates for star-forming galaxies, we use the Dale et al. (2001) and Dale & Helou (2002) SED models which are produced by a semi-empirical method to represent spectral energy distribution of star-forming galaxies in the IR region of spectrum. Dale et al. (2001) add up emission profiles of different dust families (i.e. large grains, very small grains and polycyclic aromatic hydrocarbons) which are exposed to a range of radiation field strengths in a parametrized form to generate the SED of different star-forming systems. In this way, it is possible to produce the SED corresponding to each  $R(60, 100)$  colour and scale it to the desired total infrared luminosity. In our model, we use spectral templates taken from the Dale & Helou (2002) catalog which provides 64 normalized SEDs with different  $R(60, 100)$  colours, ranging from  $-0.54$  to  $0.21$ . Some SED examples taken from this template set are illustrated in Figure 5.1.

## 5.2.4 The algorithm

Given the distribution of objects in the Universe, and their associated SEDs, it is possible to calculate the number of sources which have observed fluxes above some detection threshold,  $S_{\text{th}}$ , at a given wavelength  $\lambda = \lambda_{\text{obs}}$ :

$$N(> S_{\text{th}}) = \int \int \int Q \times \Phi(L, C, z) \frac{dV}{dz} dz dL dC, \quad (5.13)$$

where  $V$  is the volume and  $Q$  is the probability that a source with luminosity  $L$ , colour  $C$  and redshift  $z$  has an observed flux density greater than  $S_{\text{th}}$  at that wavelength (i.e., its detection probability). At each point of redshift-colour-luminosity space,  $Q$  is either 1 or 0 for any particular galaxy, but it is useful to think of  $Q$  as the average probability of detection for galaxies in each cell of that space.

In order to calculate  $Q$ , we first note that for each source with a given luminosity,  $L$ , and colour,  $C$ , there is a redshift,  $z_{\text{max}}$ , at which the observed flux is

equal to the detection threshold

$$S(L, C, \lambda_0, z_{\text{max}}) = S_{\text{th}} = \frac{(1 + z_{\text{max}})L_{\nu}(L, C, \lambda_0)}{4\pi D_L^2(z_{\text{max}})}, \quad (5.14)$$

where  $D_L$  is the luminosity distance and  $L_{\nu}$  is the rest-frame luminosity density ( $\text{W Hz}^{-1}$ ) of the object with total luminosity  $L$  and colour  $C$  at wavelength  $\lambda_0 = \lambda_{\text{obs}}/(z + 1)$ .

Now consider a cell defined by redshift interval  $\Delta z = z_2 - z_1$  (where  $z_1 < z_2$ ), luminosity interval  $\Delta L = L_2 - L_1$  and colour interval  $\Delta C = C_2 - C_1$ . Assuming negligible colour and luminosity evolution between  $z_1$  and  $z_2$ , the average detection probability in the cell is equal to the fraction of detectable objects in that cell which is:

$$Q = \begin{cases} 1 & \text{if } z_{\text{max}} > z_2 \\ \frac{D^3(z_{\text{max}}) - D^3(z_1)}{D^3(z_2) - D^3(z_1)} & \text{if } z_1 \leq z_{\text{max}} \leq z_2 \\ 0 & \text{if } z_{\text{max}} < z_1 \end{cases} \quad (5.15)$$

where  $D(z)$  is the proper distance equivalent to redshift  $z$ . For writing equation 5.15, we assumed galaxies to be distributed uniformly in space between redshifts  $z_1$  and  $z_2$ . Consequently, based on equation 5.13, the number of objects which are contributing to the source count in each cell of the redshift-colour-luminosity space is

$$\Delta N(> S_{\text{th}}) = Q \times \bar{\Phi}(L, C, z) \Delta L \Delta C \Delta V, \quad (5.16)$$

where  $\Delta V$  is the volume corresponding to the redshift interval  $\Delta z$  and  $\bar{\Phi}(L, C, z)$  is the average CLF in the cell, which for small enough  $\Delta L$ ,  $\Delta C$  and  $\Delta z$  can be written as

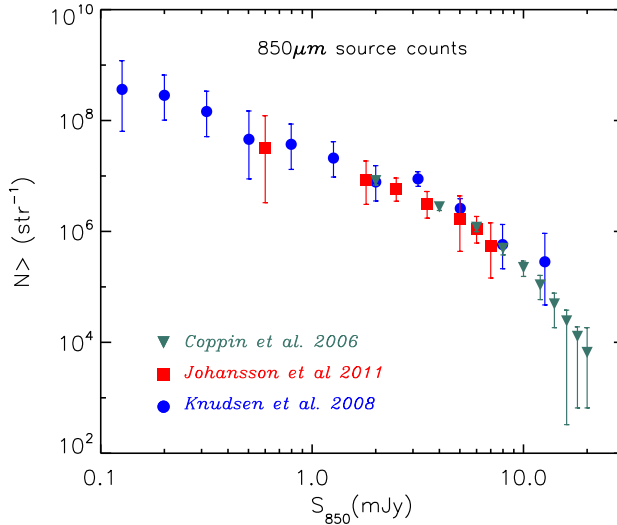
$$\bar{\Phi}(L, C, z) = \Phi(\bar{L}, \bar{C}, \bar{z}). \quad (5.17)$$

Then, the total source count is obtained by summing over the contribution from all cells (see Appendix A).

As a result of using  $Q$  and its special form as expressed in equation 5.15, our algorithm computes the source count for continuously distributed sources in the Universe, independent of the size of  $\Delta z$  which is used in equation 5.13. Thanks to this property, the computational cost required for the source count calculation is reduced significantly, and the small size of  $\Delta z$  is only necessary for accurate calculation of  $K$ -corrected SEDs and CLFs at different redshifts and not to guarantee a uniform distribution of sources (see also Appendix A).

Now that we have all the tools ready, we can test the capability of our model in reproducing the observed properties of 850  $\mu\text{m}$  sources and find which particular choices of model parameters are implied. In the following sections, we first discuss the observational constraints that we want to reproduce with our model and then we will proceed with finding the best-fit model which can reproduce those properties.





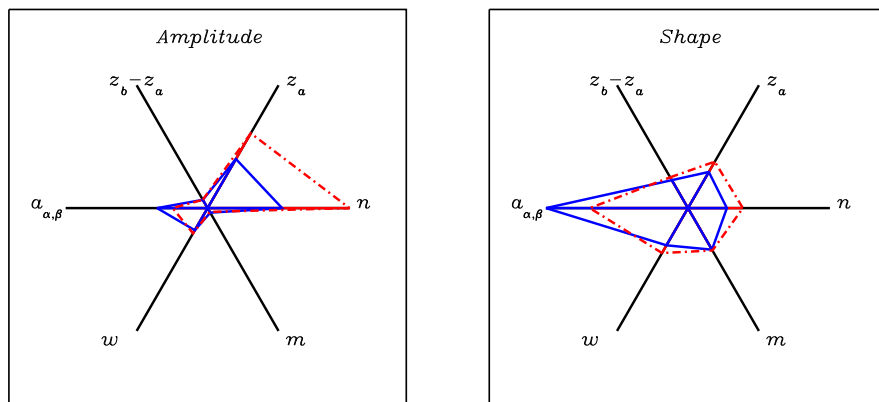
**Figure 5.2:** A compilation of some observed  $850\ \mu\text{m}$  source counts. Blue circles, green triangles and red squares are respectively data points taken from Knudsen et al. (2008), Coppin et al. (2006) and Johansson et al. (2011).

## 5.3 $850\ \mu\text{m}$ Observational Constraints

### 5.3.1 Observed $850\ \mu\text{m}$ source count

Among several existing extragalactic submm surveys which provide source counts at  $850\ \mu\text{m}$  (Coppin et al., 2006; Weiss et al., 2009; Austermann et al., 2010), the SCUBA Half-Degree Extragalactic Survey (SHADES) (Coppin et al., 2006) is the largest one which has the most complete and unbiased sample. However, this survey and other surveys which use JCMT and the same blank field method are restricted by the JCMT confusion limit of  $\sim 2\text{mJy}$  at  $850\ \mu\text{m}$  and cannot probe the source counts of the fainter population. Using a complementary method, the lensing technique has been used to probe  $850\ \mu\text{m}$  source counts to flux thresholds as low as  $0.1\text{mJy}$  (Smail et al., 2002; Knudsen et al., 2008; Johansson et al., 2011). As we will show later in this section, the sensitivity of bright and faint submm source counts to the evolution of dusty galaxies is completely different and it is essential to incorporate a large dynamic range to constrain possible evolutionary scenarios. Therefore, we use the best available observational information at both faint and bright tails of  $850\ \mu\text{m}$  source count, by combining all of the SHADES data points with those of Knudsen et al. (2008) at flux thresholds  $< 2\text{mJy}$  to assemble our reference source count which is also in agreement with other observations (see Figure 5.2).



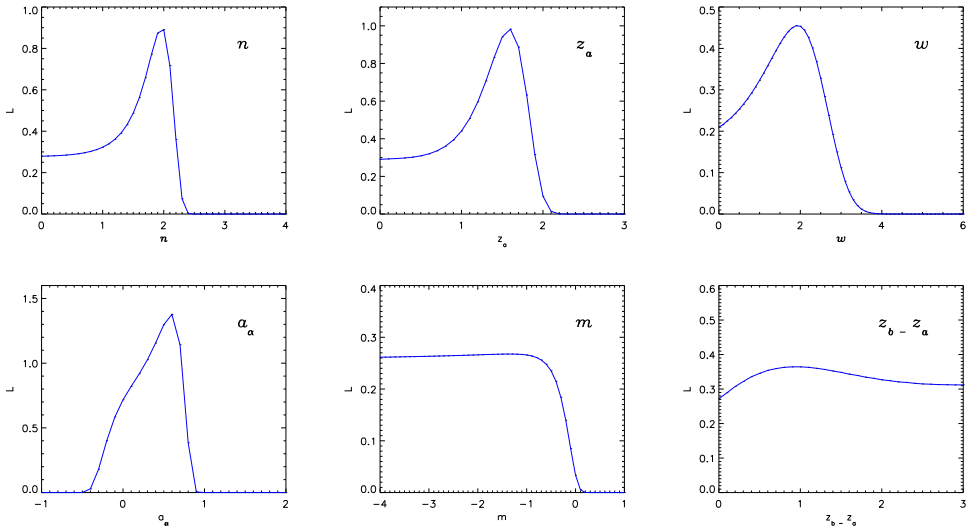


**Figure 5.3:** The sensitivity of the model to various parameters. While the left panel illustrates how much the variation of different parameters could change the amplitude of the 850  $\mu\text{m}$  source count curve, the right panel shows their effect in changing the shape of the source count curve. To measure the amplitude changes, we varied each parameter by 10% around its best-fit value and measured the difference in the total source counts. The shape sensitivity is probed by measuring the relative change each varying parameter can cause in the ratio between two source counts at faint and bright flux thresholds (i.e.  $\frac{N_{1\text{mJy}}}{N_{1\text{Jy}}}$ ). The lengths of the coloured segments on each axis shows how large those changes are. The blue (red) segments which are connected by solid (dash-dotted) lines represent the result as we increased (decreased) the value of each parameter. We show parameters  $a_\alpha$  and  $a_\beta$  on one axis since their effect in changing the source count properties is similar.

### 5.3.2 Redshift distribution of bright 850 $\mu\text{m}$ sources

Although the observed number counts at 850  $\mu\text{m}$ , especially at faint fluxes, are sensitive to the redshift distribution of infrared galaxies, they do not constrain it directly. As we will discuss later, different models with different redshift distributions can reproduce the 850  $\mu\text{m}$  source counts with the same accuracy. Therefore, it is important to impose an additional constraint on redshift distribution of those objects.

Unfortunately, studying the redshift distribution of submm galaxies is extremely difficult and there are only a few works on spectroscopically confirmed redshift distribution of bright submm galaxies with observed fluxes greater than  $\sim 4 - 5$  mJy (Chapman et al., 2005; Wardlow et al., 2011). Those studies show a redshift distribution peaking around  $z \sim 2$  (see the right panel in Figure 5.6). We use this redshift distribution as one of our observed constraints and force our model to reproduce a redshift distribution which peaks at the same redshift.



**Figure 5.4:** The likelihood (i.e.  $\exp(-\frac{1}{2}\chi^2)$ ) distribution of different parameters for the best-fit model. For each panel, we fixed all the parameters at their best-fit values and changed only one parameter at a time. By varying each parameter, the quality of the fit and hence the likelihood is changing. We normalized the amplitude of curves such that the area under each curve (i.e. the total probability) be unity.

## 5.4 Finding 850 $\mu\text{m}$ best-fit Model

In order to be able to extract meaningful trends and differentiate between them, it is important to keep our phenomenological model as simple as possible. It is therefore desirable to find the minimum number of free parameters without which the model cannot produce an acceptable result. Furthermore, the role of different parameters are not identical: while some parameters are not strongly constrained, small changes in others can change the result significantly. It is also important to look at degeneracies between various parameters; some parameters are not independent and varying one may be compensated by varying the others. In this section we investigate our model to understand those issues before presenting our best-fit model.

### 5.4.1 The source count curve: Amplitude vs. Shape

To match the observed source counts, our model should be able to reproduce both the typical number of sources (i.e., the amplitude of the source count curve as a function of flux threshold) and the ratio between the source counts at faint and bright flux thresholds (i.e., the shape of the source counts curve). Following this line of argument, we can categorize our model parameters into two different

groups: those which play a stronger role in forming the amplitude of the curve and those which mainly affect its shape. As mentioned earlier, the luminosity evolution (see equation 5.7) has the dominant role in determining the amplitude of the source count curve while its shape is controlled by other ingredients of our model which are colour evolution and the evolution of LF slopes.

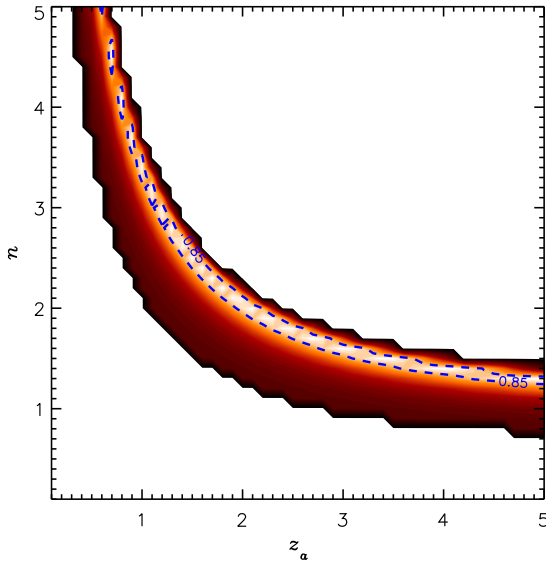
These trends are illustrated in Figure 5.3. In the left panel, the relative contribution of each parameter in changing the amplitude of the source count curve is visualized; the length of the coloured segments along each axis (coloured segments along different axis are connected to each other) represents the sensitivity of source count amplitude to that parameter. The length of coloured segments is computed by changing all the model parameters (one at a time) by the same fraction, for instance 10%, and measure the change caused in the total source count. The segments which are connected by dot-dashed (red) and solid (blue) lines correspond respectively to a decreasing and an increasing parameter. In the right panel, the relative contribution of different parameters in controlling the shape of the source counts curve is shown. In this figure, the length of coloured segments represents the change in the ratio between faint and bright source counts. This time, we measured how much the ratio between a very faint source count, like 1 mJy, and a very bright one, like 1 Jy, is changing due to a fixed change in different parameters (e.g., 10%). The dot-dashed (red) and solid (blue) lines connect the coloured segments which correspond to an increase or a decrease in the values of the parameters.

As is evident from these diagrams, the amplitude of the source count curve is strongly sensitive to the parameters of the luminosity evolution, particularly  $z_a$  and  $n$ , while its shape is determined mainly by an evolution in the shape of LF and/or the strength of the colour evolution.

## 5.4.2 The luminosity evolution

The luminosity evolution is the backbone of our model and has the most important role in producing the observed source counts and the redshift distribution of 850  $\mu\text{m}$  objects. However, it is not surprising that the calculated source count is not strongly sensitive to the model properties at high redshifts. This is mainly because objects at very high redshifts have decreasing fluxes (despite the advantageous  $K$ -correction). Consequently, the chosen value of  $m$  has a negligible effect on the amplitude and shape of the source count curve (see Figure 5.3). Nevertheless, one should note that this argument would not necessarily work if the population of bright IR galaxies continued to evolve with look-back time for all redshifts. In other words, the exact value of  $m$  is not strongly constrained by the observed source counts and all the negative values can produce similar results (see the middle panel in the bottom row of Figure 5.4).

Contrary to  $m$ , the source count curve is very sensitive to the exact values of  $n$  and  $z_a$  which respectively control the rate by which the characteristic luminosity of IR galaxies increases with redshift and up to which redshift this



**Figure 5.5:** Map of likelihoods for models with different  $n$  and  $z_a$  where  $[w, a_\alpha, a_\beta] = [0, 0, 0]$ . The region which contains the best-fit models is in the middle of the coloured band and surrounded by the dashed blue contours which indicate the likelihood of 0.85.

growth continues. However, there are two solutions for reproducing the same source counts: one is to increase the characteristic IR luminosity rapidly up to a relatively low redshift and the other one is to increase the characteristic IR luminosity by a moderate rate but for a longer period of time (i.e., up to higher redshifts). Therefore, as illustrated in Figure 5.5, there is a degeneracy between  $n$  and  $z_a$  and one cannot constrain them individually by looking at the observed source counts. However, additional information about the redshift distribution of submm galaxies or the slope by which the characteristic IR luminosity is growing at low redshifts could resolve this degeneracy. As we will discuss later, the former constraint is used in finding our best-fit model.

Unlike  $n$  and  $z_a$ , the length of the redshift interval during which the luminosity evolution remains constant before starting to decline,  $z_b - z_a$ , is not an essential part of our model. We only introduced this feature to have a smoother transition between growing and declining characteristic IR luminosity and also redshift distribution of submm galaxies. As Figures 5.3 and 5.4 show, variations in the adopted value for  $z_b - z_a$  do not change the source count significantly. Moreover, any change in its value could be compensated by a very small change

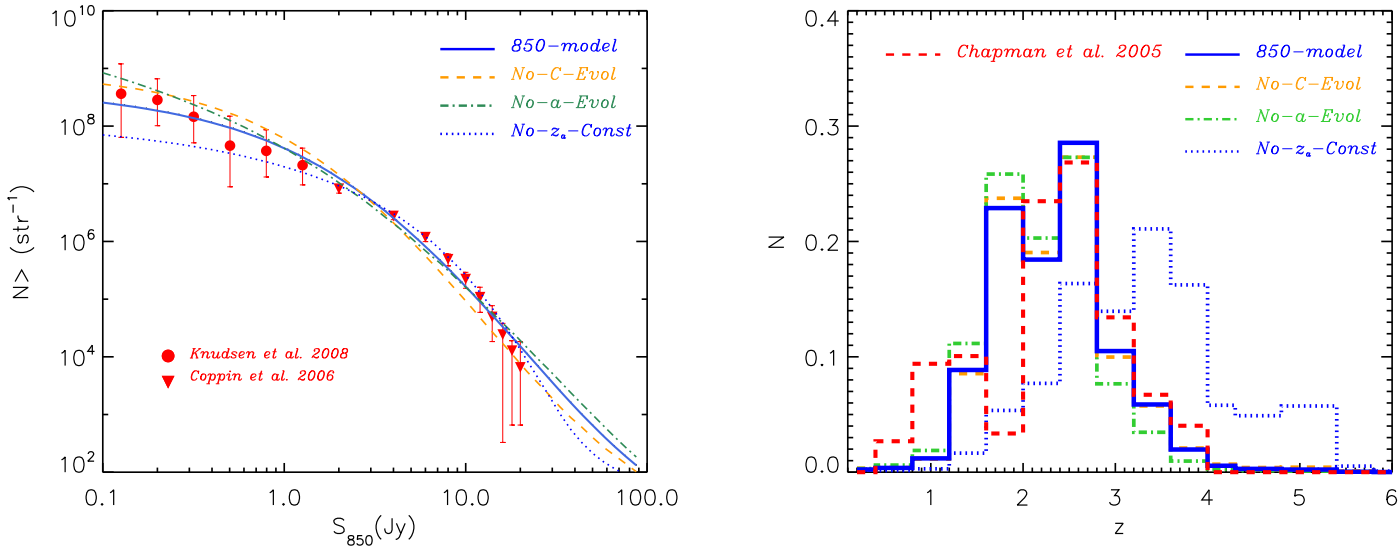
in  $n$  and/or  $z_a$ .

Based on these considerations and as we already mentioned in §5.2.2, we reduce the number of free parameters we use in our model by choosing  $m = -1$  and  $z_b - z_a = 1$ . Furthermore, the observed redshift distribution of submm galaxies which peaks around  $z \sim 2$  (Chapman et al., 2005; Wardlow et al., 2011), limits the acceptable values of  $z_a$  to  $\sim 1.6$  (see the right panel in Figure 5.6).

## 5.5 Other necessary model ingredients

Although the luminosity evolution is necessary for producing correct number of observable sources, it is not sufficient to provide a correct shape for the source count curve. Based on our experiments by varying different parameters of the luminosity evolution, one can either get a good fit at the faint number counts and under-produce the bright end or produce correctly the bright source counts and over-estimate the faint objects. In other words, other model ingredients like the colour evolution and/or the evolution of LF slopes are required to adjust the shape of the source count curve appropriately in order to fit the faint and bright source counts at the same time. For instance, the colour evolution can be used to help the model with under-production of bright sources and the evolution of LF slopes can compensate for the over-production of the faint sources.

As a starting point, we have first explored fits using either the colour evolution or the evolution of the LF slopes, in order to keep the model as simple as possible. We did this by trying to fit the 850  $\mu\text{m}$  source counts once using a combination of the luminosity evolution together with the colour evolution when the slopes of LF do not evolve (i.e., No-a-Evol) and once using the luminosity evolution combined with the evolving LF slopes, without evolving the colour distribution (i.e. No-C-Evol). The results are shown in Figure 5.6 and are compared with the best-fit result when the colour evolution and the evolution of LF slopes are both active (i.e., 850-model). The quality of the fit for “850-model” is better than both of the other cases which may be attributed to the additional free parameter used in the “850-model”. However, both “No-a-Evol” and “No-C-Evol” models have unfavourable implications. The best-fit for the “No-C-Evol” case (see Table 5.1) requires a large increasing slope in the luminosity evolution function which gives rise to a violation of the counts of IR sources at short wavelengths (e.g. at 70  $\mu\text{m}$ , see §5.6). The “No-a-Evol” fit on the other hand, requires a steep colour evolution which is too extreme to be acceptable, since it would imply that at redshifts  $z > 1$  all infrared sources, independent of their luminosities, have the same dust temperatures as low as  $T \sim 20\text{K}$ . However, when both colour evolution and evolution of the bright and faint-end slopes of the LF are allowed simultaneously, these problems disappear which makes this solution preferable.



**Figure 5.6:** Left panel: The best-fit model which is constrained by  $850\ \mu\text{m}$  source count and the redshift distribution of submm galaxies (the blue solid line) is illustrated next to the reference observed data points (see §5.3). For comparison, three other best-fit models are also plotted: the best-fit model without colour evolution shown by the dashed orange line (i.e., “No-C-Evol”), a model without evolving LF slopes shown by the dot-dashed green curve (i.e., “No-a-Evol”) and finally the best-fit model which is only constrained by the source count is shown using the dotted blue curve. Right panel: the comparison between the observed redshift distribution of  $850\ \mu\text{m}$  galaxies which are brighter than 5 mJy (Chapman et al., 2005) and what our best-fit model implies. The histogram shown in red with dashed line is the observed probability distribution and the histogram with solid blue line shows the probability distribution of similar objects in our model. For comparison, the best-fit model without constrained redshift distribution, is shown with dotted blue histogram (“No-z<sub>a</sub>-Const”). The other two models, “No-C-Evol” and “No-a-Evol” are respectively shown by dashed orange and dot-dashed green lines.

**Table 5.1:** Parameters which define different best-fit models constrained to reproduce the observed source counts at  $850\ \mu\text{m}$ . All the models are using  $z_b - z_a = 1$  and  $m = -1$  and except the "No- $z_a$ -Const" model, all of them are constrained to reproduce the redshift distribution of submm galaxies and therefore use  $z_a = 1.6$ . The predicted source count each of those models and their implied redshift distribution for submm sources is illustrated in Figure 5.6

Model	$n$	$z_a$	$w$	$a_\alpha$	$a_\beta$
850-model	2.0	1.6	2.0	0.6	0.4
No-C-Evol	2.8	1.6	0	0.6	0.6
No-a-Evol	1.6	1.6	5.6	0	0
No- $z_a$ -Const	2.4	3.6	2.4	1.0	2.2

### 5.5.1 The $850\ \mu\text{m}$ best-fit model

As we showed, it is necessary to have an evolutionary model which incorporates a luminosity evolution, colour evolution and also evolving LF slopes, in order to fit the observed  $850\ \mu\text{m}$  source counts. We also argued that we can fix some of the initial parameters of the model since they have no significant effect on the results and chose proper values for them (i.e.,  $m = -1$  and  $z_b - z_a = 1$ ). Moreover, we showed that we need to set the redshift at which the luminosity evolution peaks to  $z_a = 1.6$  to reproduce the redshift distribution of submm galaxies correctly which also resolves the degeneracy between  $z_a$  and  $n$ . After taking into account all of those considerations, we end up with 4 free parameters in our model which are needed to be adjusted properly to reproduce our observed sample of  $850\ \mu\text{m}$  source counts; those parameters are the slope of the colour evolution,  $w$ , the rate by which the faint and bright end slope of LF is changing with redshift,  $a_\alpha$  and  $a_\beta$ , and finally the growth rate of the characteristic IR luminosity,  $n$ .

Since our algorithm for calculating the source count is fast and accurate, contrary to much more cumbersome Monte-Carlo-based approaches, we can perform a comprehensive search in the parameter space for the best-fit model instead of choosing it "by hand". We split each dimension of relevant regions of parameter space into equally spaced grids and calculate the source count for modes associated with each node of our grid structure. Then we calculate the likelihood of each model for reproducing the observed source counts (i.e.  $\exp(-\frac{1}{2}\chi^2)$ ). Finally, we choose the model with maximum likelihood as our best-fit model. The parameters which define the best-fit model for  $850\ \mu\text{m}$  source counts, "850-model", is shown in Table 5.1 together with those of "No-C-Evol" and "No-a-Evol" which we discussed in the previous section. Those models are also compared with observational data sets used for constraining them, in Figure 5.6 where also the redshift distribution implied by each model is illustrated (in the right panel).

It is also interesting to inspect the properties of a best-fit model in which the

peak of the luminosity evolution,  $z_a$ , is not fixed. The parameters which define this model are presented in the last row of Table 5.1 and its source count and redshift distribution are shown by the blue dotted lines in Figure 5.6. The quality of the fit for this model is even better than “850-model” except at very low flux thresholds (unsurprisingly, since it has an additional free parameter, namely  $z_a$ ). The rising slope of luminosity evolution,  $n$ , and the colour evolution,  $w$ , in this model are not drastically different from the “850-model” but its peak of luminosity evolution happens at higher redshifts and the model requires much steeper slope evolutions to compensate for too many observable sources. Not only is this model unable to produce the observed redshift distribution (see the right panel in Figure 5.6), it predicts too few observable sources at shorter wavelengths which makes it unfavorable.

## 5.6 Other wavelengths

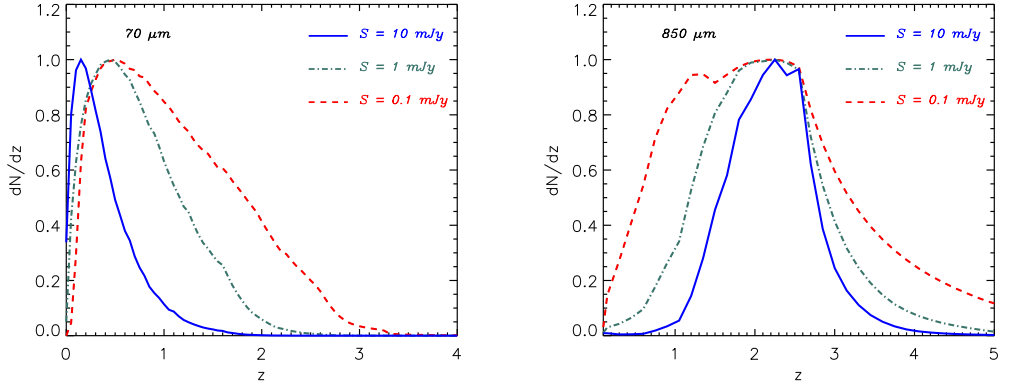
In the previous section, we discussed our model parameters and their role in producing the observed 850  $\mu\text{m}$  source counts and the redshift distribution of submm galaxies. We used those observed quantities as constraints to find our best-fit model, “850-model”. If we assume the evolution scenario that our best-fit model suggests is correct, and also the set of SEDs we used for reproducing the 850  $\mu\text{m}$  source count are good representatives for real galaxies, then we expect the same model, together with the same set of SEDs, to reproduce the observed number counts of IR sources at other wavelengths. Moreover, one can use the source count at other wavelengths to find a best-fit model for that specific wavelength and again expects to reproduce the source count at other wavelengths correctly.

Before examining those expectations, it is important to note that there are some fundamental differences between the source counts at shorter wavelengths and in the submm. Most importantly, as illustrated in Figures 5.7 and 5.11, objects observed at short wavelengths (e.g., 70  $\mu\text{m}$ ) are mainly at low redshifts while the observed submm galaxies are distributed in a wider redshift interval and at higher typical redshifts. Also, the brightest objects at 70  $\mu\text{m}$  are the closest ones which is not the case for brightest submm galaxies<sup>1</sup> (Berta et al., 2011). Moreover, there are additional physical processes, like AGNs, which can change the energy output of galaxies at shorter wavelengths. Since we do not take into account AGNs as a separate population, our model is not expected to reproduce necessarily good results for wavelengths shorter than 60 – 70  $\mu\text{m}$  (see also the discussion in §5.2.3). Going from longer to shorter wavelengths, the observed sources are typically at lower and lower redshifts. Consequently, the source counts at short wavelengths are only sensitive to the very low redshift

---

<sup>1</sup>In fact the wide distribution of observed submm galaxies in redshift space, combined with the availability of a measured redshift distribution are the primary reasons we chose their observational properties to constrain our model.





**Figure 5.7:** The modeled redshift distribution of objects which are observed with different flux thresholds at  $70 \mu\text{m}$  (in the left panel) and  $850 \mu\text{m}$  (in the right panel). Three different flux thresholds which are 0.1, 1 and 10mJy are shown respectively using solid (blue), dot-dashed (green) and dashed (red) lines.

**Table 5.2:** Parameters which define different best-fit models constrained to reproduce the observed source counts at different wavelengths. The first column,  $\lambda$ , indicates the wavelength for which the model is constrained to produce the best fit to the observed source counts. All the models are forced to reproduce the redshift distribution of submm galaxies and therefore use  $z_a = 1.6$ ,  $z_b - z_a = 1$  and  $m = -1$ . The predicted source count each of those models is implying for different wavelengths is illustrated in Figure 5.8 and Figure 5.9

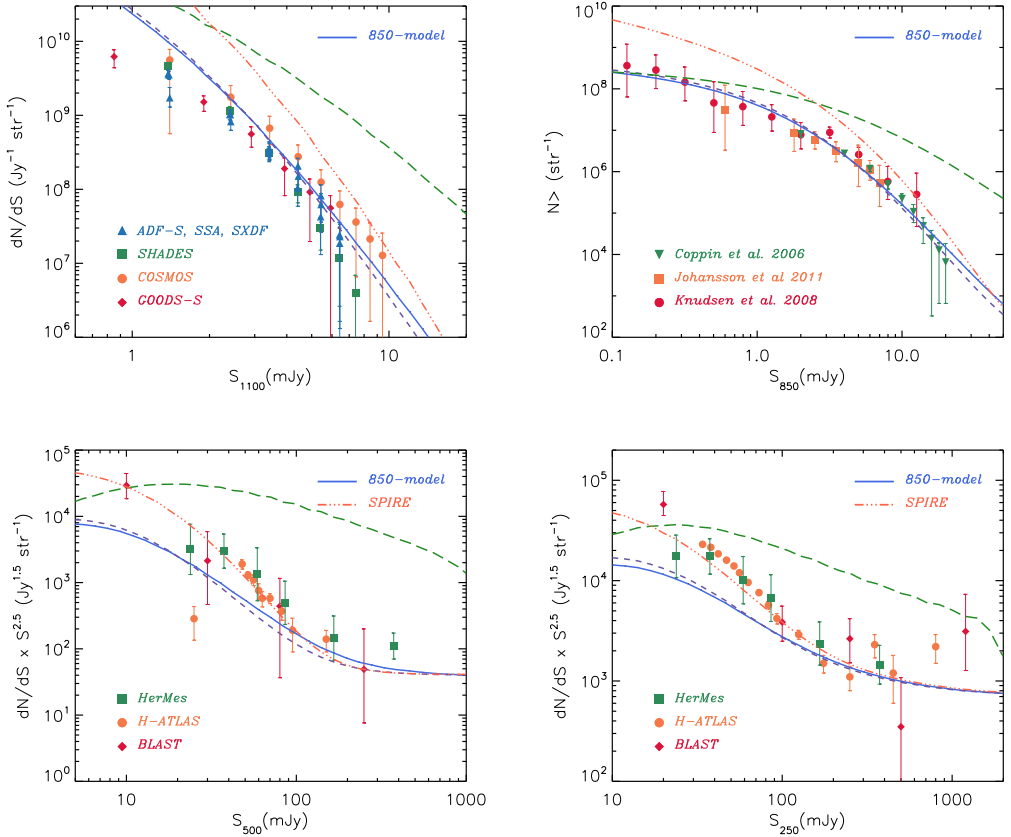
$\lambda$	$n$	$w$	$a_\alpha$	$a_\beta$
$850 \mu\text{m}$	2.0	2.0	0.6	0.4
$500 \mu\text{m}$	2.6	3.8	0.0	0.6
$350 \mu\text{m}$	3.0	4.4	-0.2	1.4
$250 \mu\text{m}$	3.0	4.4	0.0	1.4
$160 \mu\text{m}$	2.2	4.2	0.8	0.2
$70 \mu\text{m}$	2.2	2.0	0.6	0.6

properties of our model. In fact, the  $70\ \mu\text{m}$  source count is mainly sensitive to the growth rate of the luminosity evolution,  $n$ , and the evolution in LF slopes; but by going to longer wavelengths, the colour evolution and the redshift at which the luminosity evolution is peaking,  $z_a$ , become more important. In other words, if a model reproduces the observed source counts at short wavelengths, this forms a confirmation that the evolution at low redshifts is represented correctly but for a best-fit model constrained by observations at shorter wavelengths, the properties of model at intermediate and high redshifts are not strongly constrained.

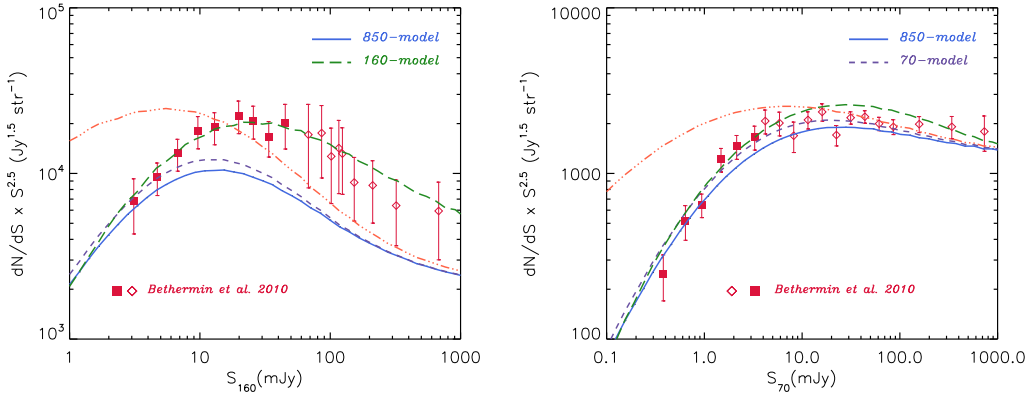
To investigate the performance of different best-fit models constrained by source count observations of wavelengths other than  $850\ \mu\text{m}$ , we use the same procedure we incorporated in finding the "850-model" and only vary the effective parameters of the model, namely  $n$ ,  $w$ ,  $a_\alpha$  and  $a_\beta$ , to fit the observed data. The parameters which define best-fit models at different wavelengths are shown in Table 5.2 and the source counts they produce at different wavelengths are illustrated next to the observational data points in Figure 5.8 and Figure 5.9. Different panels are for source count at different wavelengths ranging from  $1100\ \mu\text{m}$  on the top left to  $70\ \mu\text{m}$  on the bottom right. We do not show the results for  $350\ \mu\text{m}$  to maintain the symmetry of figures since for this wavelength models and their comparison with observed data are in many respects identical to the case of  $250\ \mu\text{m}$ . In Figure 5.8 and Figure 5.9, the source counts produced by different models are also shown using lines with different styles and colours: purple short-dashed for  $70\ \mu\text{m}$ , green long-dashed for  $160\ \mu\text{m}$ , orange dot-dot-dashed for  $500\ \mu\text{m}$  and finally solid blue lines for  $850\ \mu\text{m}$  (i.e the "850-model"). In the following we discuss those results by categorizing them in different wavelength ranges, namely  $850\ \mu\text{m}$  and  $1100\ \mu\text{m}$  as long submm wavelengths,  $500\ \mu\text{m}$ ,  $350\ \mu\text{m}$  and  $250\ \mu\text{m}$  as SPIRE or intermediate wavelengths and finally  $70\ \mu\text{m}$  and  $160\ \mu\text{m}$  as short wavelengths.

### 5.6.1 Long submm wavelengths: $850\ \mu\text{m}$ and $1100\ \mu\text{m}$

As we mentioned earlier, the SED of star forming galaxies at long submm ranges is essentially controlled by the Rayleigh-Jeans tail of the dust emission which simply falls off smoothly (see Figure 5.1). This means that if a model reproduces the observed counts at  $850\ \mu\text{m}$ , a good fit to the observed counts at similar and longer wavelengths is guaranteed that given the distribution of sources which produce those counts have a similar redshift distributions, which in turn makes those counts equally sensitive to different parameters in our model. As illustrated in the top panels of Figure 5.8, both models which are constrained by observed  $70\ \mu\text{m}$  and  $850\ \mu\text{m}$  counts and produce a good fit to  $850\ \mu\text{m}$  data, also produce a good fit to  $1100\ \mu\text{m}$  source counts. Our experiments with other models also confirm that the quality of the fit they produce for observed counts at  $850\ \mu\text{m}$  and  $1100\ \mu\text{m}$  is highly correlated. Both of the models constrained by  $160\ \mu\text{m}$  and SPIRE (e.g.  $500\ \mu\text{m}$ ) source counts over produce the long submm wavelength source counts: the "160-model" over-produces



**Figure 5.8:** In different panels, the best-fit model predictions for cumulative source counts at 850  $\mu\text{m}$  and differential source counts at 1100, 500 and 250  $\mu\text{m}$  are plotted next to the observed data (see also the same for 160 and 70  $\mu\text{m}$  in Figure 5.9). The 1100  $\mu\text{m}$  observational data sets are from Scott et al. (2010) (red diamonds) Austermann et al. (2009, 2010) (orange circles and green squares respectively) and Hatsukade et al. (2011) (blue triangles); data at 850  $\mu\text{m}$  is from Coppin et al. (2006) (green triangles), Knudsen et al. (2008) (red circles) and Johansson et al. (2011) (orange squares). The data for 500 and 250  $\mu\text{m}$  source counts are from the BLAST experiment taken from Patanchon et al. (2009) (red diamonds), Herschel data taken from Oliver et al. (2010) (green squares) and Clements et al. (2010) (orange circles). We also used one data point at 350  $\mu\text{m}$  from the SHARC2 survey (green triangle) (Khan et al., 2007). Different lines correspond to the source counts produced by various models which are constrained to fit the observed source counts at different wavelengths: blue solid line shows the 850-model, purple dashed line shows the "70-model" and green long-dashed and orange dot-dot-dashed lines are for "160-model" and "500-model" respectively (see Table 5.2)



**Figure 5.9:** The same as Figure 5.8 for 160 and 70  $\mu\text{m}$ . The data at 160 and 70  $\mu\text{m}$  are based on Spitzer observations taken from Bethermin et al. (2010) where the data points shown by filled squares represent stacking results.

submm source counts mainly at bright flux thresholds due to its extreme colour evolution (see Table 5.2) but fits the observations at faint fluxes, thanks to its luminosity evolution which is similar to those of "850-model" and "70-model". However, SPIRE constrained models (e.g. "500-model") have stronger evolutions both in terms of luminosity and colour and over-produce the data in all observed fluxes.

One should note that the 1100  $\mu\text{m}$  data points are highly incomplete below 3mJy (Hatsukade et al., 2011) and are prone to large field to field variations and relatively large errors in the whole range of observed fluxes. Moreover, the observed counts are at bright flux thresholds which makes them sensitive to a narrower redshift interval in comparison to the fainter flux thresholds (see the right panel of Figure 5.7) and because of those reasons they cannot constrain models better than what 850  $\mu\text{m}$  source counts are capable of. Therefore, we only consider the "850-model" as a model constrained by long submm source counts.

### 5.6.2 SPIRE intermediate wavelengths: 500 $\mu\text{m}$ , 350 $\mu\text{m}$ and 250 $\mu\text{m}$

The three 500  $\mu\text{m}$ , 350  $\mu\text{m}$  and 250  $\mu\text{m}$  wavelengths are close together and besides being in the middle of wavelength ranges we study, they have intermediate properties with respect to the redshift range each wavelength is mostly sensitive to: as it is shown in Figure 5.11, while at the bright flux thresholds the source counts mainly consist of low redshift sources, at fainter fluxes they are sensitive to the intermediate (i.e.  $1 < z < 2$ ) and high redshifts (i.e.  $2 < z$ ). Natur-

ally the general behavior of models for  $500\ \mu\text{m}$  is closer to  $850\ \mu\text{m}$  while  $250\ \mu\text{m}$  is close to shorter wavelengths. However, those intermediate wavelengths are closely similar to each other more than being similar to other wavelengths. Consequently, the best-fit models constrained by SPIRE wavelengths have similar parameters (250-model and 350-model have almost identical parameters) and any of those models agrees with the observed source counts of the other two. However, SPIRE-constrained models all require steep luminosity evolution and strong colour evolution in addition to a steepening bright end slope of the LF with redshift (i.e. positive  $a_\alpha$  together with almost zero  $a_\beta$ , see Table 5.2). The first two mechanisms produce too many objects in comparison to what is needed for the source counts at other wavelengths while the steeper bright end of the LF partially compensate the over-production of bright objects: as is shown in Figure 5.8 and Figure 5.9, SPIRE-constrained models over-produce the faint source counts both at longer and shorter wavelengths but the steep bright end of LF produces results which are close to the observed faint source counts at very long wavelengths (i.e.  $850\ \mu\text{m}$  and  $1100\ \mu\text{m}$ ) which in turn is too strong to leave enough sources required for the bright  $160\ \mu\text{m}$  sources. As we will discuss later, it is not surprising that SPIRE-models agree with bright  $70\ \mu\text{m}$  counts since they are essentially  $z \sim 0$  objects for which evolutionary mechanisms in the model barely have any effect.

While the SPIRE models fail to agree with observations at shorter and longer wavelengths, the two models which are successful at long submm ranges (i.e. the "850-model" and "70-model") also have a reasonable agreement with the SPIRE data. However, they under-produce the counts at  $S_{th} < 100\text{mJy}$  by a factor of  $\sim 2$ . As we will show later, assuming the "850-model" to be correct, this under-production is a hint for a population of cold luminous IR galaxies residing only in low and intermediate redshifts (i.e.  $z \sim 1$ ). This is also consistent with the steep colour evolution which is implied by best-fit models at those wavelengths which produces enough cold galaxies at lower redshifts.

### 5.6.3 Short wavelengths: $160\ \mu\text{m}$ and $70\ \mu\text{m}$

At short wavelengths like  $70\ \mu\text{m}$ , the K-correction is not strong enough to counteract the effect of cosmological dimming which makes the source counts at these wavelengths almost insensitive to the properties of IR galaxies at high redshift. Since the starting point of our model is the observed distribution of IRAS galaxies which are selected to be local galaxies with  $S_{60\ \mu\text{m}} > 1\text{Jy}$ , and the SED templates we use are extensively tested to match these galaxies, any variation of our model by construction cannot disagree with bright  $70\ \mu\text{m}$  counts<sup>2</sup>. However, at fainter flux thresholds (e.g.  $S_{th} < 100\text{mJy}$ ) the sensitivity of  $70\ \mu\text{m}$  counts to low and intermediate redshifts,  $0.5 < z < 2$ , increases (see Figure 5.11) which

<sup>2</sup>Unless the luminosity of IR galaxies increase with redshift extremely steep to produce high- $z$  objects which are observed in  $70\ \mu\text{m}$  band as bright as the brightest local IR galaxies despite the cosmological dimming and negative K-correction.

makes them more sensitive to the slope of the luminosity evolution. As a result, all the models agree with  $70\ \mu\text{m}$  observations at bright flux threshold while the SPIRE-model which has steeper luminosity evolution (i.e. greater  $n$ , see Table 5.2) diverges from observed counts and over-produces them. It should be noted that the "70-model" which is tuned to agree with mainly low- $z$  IR sources performs as good as "850-model" in longer wavelengths.

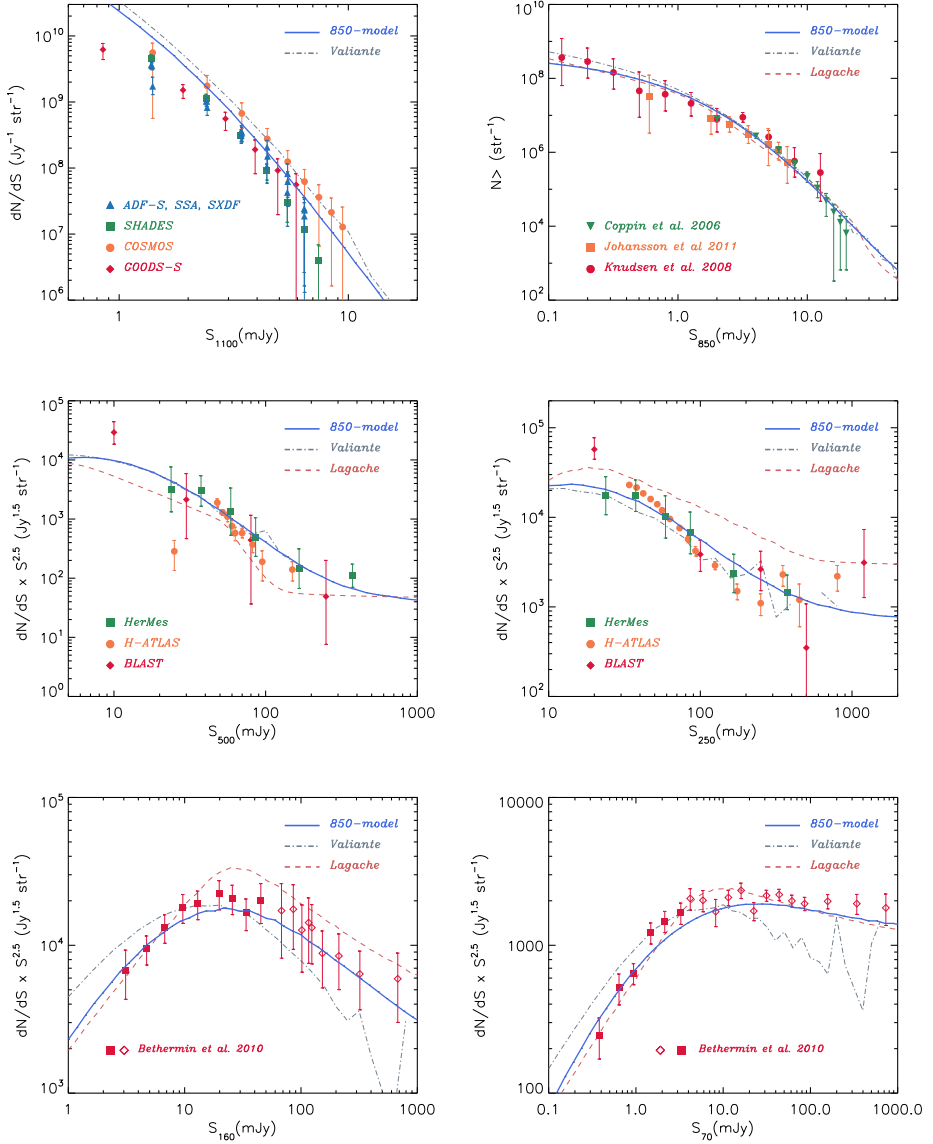
At  $160\ \mu\text{m}$ , the redshift distribution of sources is almost identical to the distribution of galaxies which are responsible for  $70\ \mu\text{m}$ : the bright end of the counts is governed by local IR galaxies while for flux thresholds  $S_{th} < 100\text{mJy}$  the intermediate redshifts (i.e.  $0.5 < z < 2$ ) become important. However, the flux threshold at which the turn over between domination of local and farther sources happens is one order of magnitude higher for  $160\ \mu\text{m}$  sources (see bottom row in Figure 5.11). Moreover, at  $160\ \mu\text{m}$  the domination of intermediate redshift sources in shaping the faint source counts is slightly stronger than at  $70\ \mu\text{m}$ . In other words, the  $160\ \mu\text{m}$  source counts are slightly more sensitive to the distribution and evolution of distant IR galaxies in comparison to  $70\ \mu\text{m}$  counts.

The best-fit model which is constrained by  $160\ \mu\text{m}$  source counts has similar slope of luminosity evolution,  $n$ , compared with the "850-model" and "70-model" but requires much stronger colour evolution and larger fraction of bright objects (i.e. steeper faint end slope and flatter bright end slope of LF at higher redshifts).

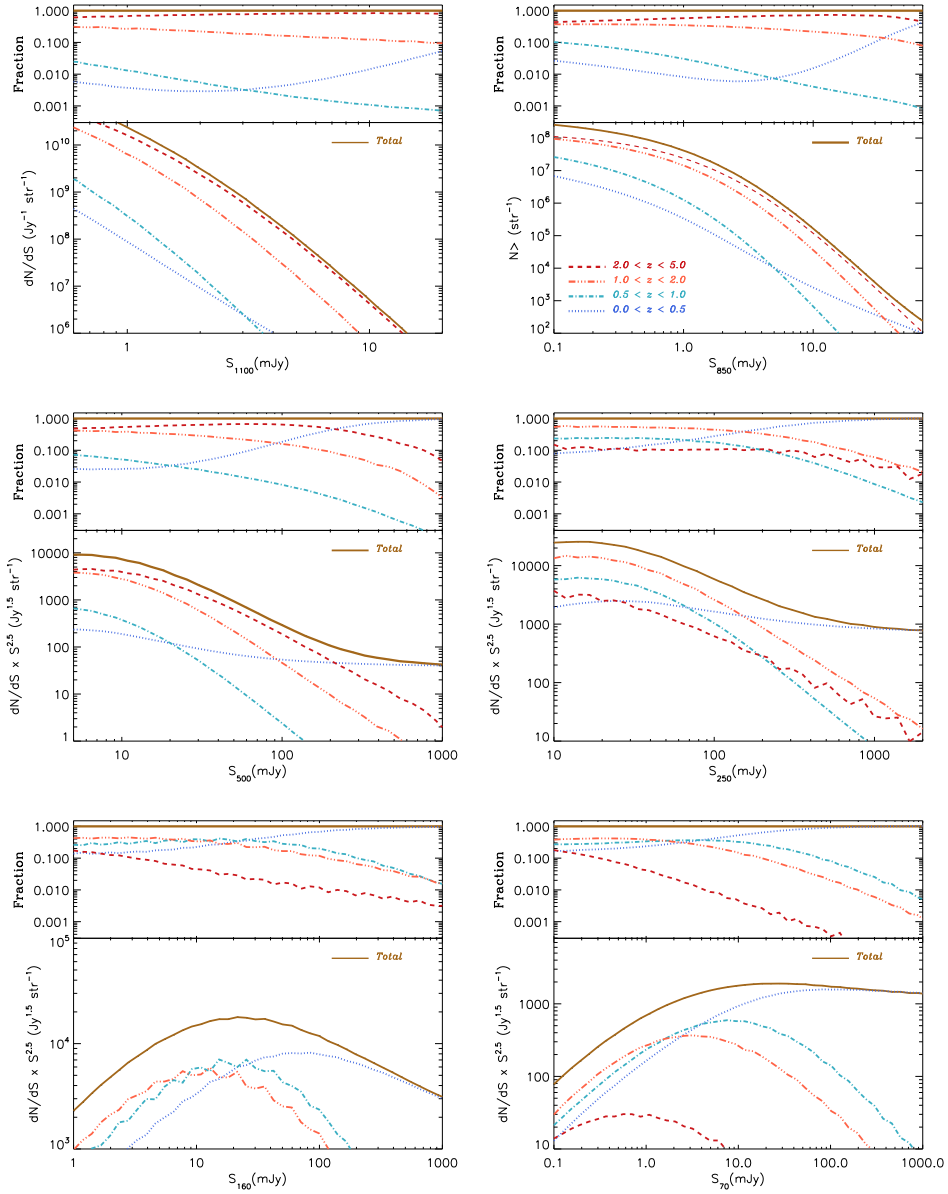
The SPIRE-models on the other hand, violate the observed  $160\ \mu\text{m}$  counts by under-producing them at bright end and over-producing them at faint fluxes. The "850-model" and "70-model" on the other hand, match the faintest  $160\ \mu\text{m}$  count but under produce the brighter objects by a factor of  $\sim 2$  which is the same factor showing up in the difference between what those two models produce and observed faint SPIRE counts. A similar discrepancy between models and  $160\ \mu\text{m}$  counts has been pointed out in some recent works (Le Borgne et al., 2009; Valiante et al., 2009); while the Le Borgne et al. (2009) best-fit model which produces correct source counts at  $70\ \mu\text{m}$  and  $850\ \mu\text{m}$  deviates from observations between  $10 < S_{160} < 100\ \text{mJy}$  by a factor of  $\sim 2$ , the Valiante et al. (2009) model underestimates number counts at  $160\ \mu\text{m}$  by a factor of  $\sim 5$ . We discuss this issue further in §5.6.4.

### 5.6.4 A best-fit model for all wavelengths

As we showed in previous subsections, the required evolution scenarios for different models which are constrained by various wavelengths is too diverse to be reconciled in a single model; however, the  $70\ \mu\text{m}$  and  $850\ \mu\text{m}$  source counts can be explained by the same model, either the "850-model" or "70-model" which are almost identical in terms of implied evolution and distribution of IR galaxies and predicted source counts at different wavelengths. While  $70\ \mu\text{m}$  source counts are produced mainly by local and low redshift objects,  $850\ \mu\text{m}$  sources are distributed in higher redshifts and in a wide redshift interval. This means the model which can explain the source count at both  $70\ \mu\text{m}$  and  $850\ \mu\text{m}$  gives a



**Figure 5.10:** The same set of observed source counts which is shown in Figure 5.8 and Figure 5.9 is also illustrated here. The blue solid line shows the result of "850-model" using a modified set of SED (see §5.6.4) to correct under-production of bright 160  $\mu\text{m}$  sources and faint SPIRE source. Source count produced by two different models, Lagache et al. (2004) and Valiante et al. (2009), are also respectively shown by orange dashed line and gray dot-dashed lines. While Valiante et al. (2009) and "850-model" counts at 160  $\mu\text{m}$  are almost identical, we did not find Lagache et al. (2004) model predictions at 1100  $\mu\text{m}$  to include them in the top-left panel.



**Figure 5.11:** The fractional contribution of sources in different redshift bins in producing the observed source counts at different wavelengths are illustrated. The blue Dotted, cyan dot-dashed, orange dot-dot-dashed and red dashed lines respectively show the source count produced by objects in  $0 < z < 0.5$ ,  $0.5 < z < 1$ ,  $1 < z < 2$  and  $2 < z < 5$  redshift intervals while the total source count is shown using the brown solid line. In the top section of each panel, the fraction of sources for each redshift bin in producing source counts at different flux thresholds is plotted.



consistent distribution of IR galaxies from local Universe to very high redshifts. In other words, the same evolution scenario for star-forming galaxies which we observe locally, can also produce the observed  $850\ \mu\text{m}$  counts. This scenario, contrary to what is implied by the best-fit models constrained at other intermediate wavelengths, does not strongly violate observed source counts at other bands, though it under-produces them at some flux thresholds. On the other hand, models which agree with observations at intermediate wavelengths require extreme color evolutions to produce larger number of cold objects in intermediate redshift ranges which is relevant for the source count at those wavelengths (i.e.  $1 < z < 2$ , see the left panel in the middle of Figure 5.11). However, this steep colour evolution at higher redshifts (i.e.  $z > 2$ ) produces too many observable sources at  $850\ \mu\text{m}$ .

Moreover, as we mentioned earlier different authors with fundamentally different models have reported inconsistencies in their best-fit models (which fit the long and short wavelengths) and intermediate source counts (Le Borgne et al., 2009; Valiante et al., 2009)<sup>3</sup> and the underestimation of  $160\ \mu\text{m}$  (and SPIRE wavelengths) seems to be model independent.

There are two main possibilities which can explain the origin of discrepancy between the "850-model" and source counts at SPIRE band and  $160\ \mu\text{m}$ , besides doubting the validity of observed counts: (i) the IR SEDs in our model are not representative and should be modified in a certain way to produce more flux at observed wavelengths  $\lambda_{obs} \sim 160 - 500\ \mu\text{m}$  or (ii) existence of a population of cold galaxies at relatively low redshifts. Indeed, there is a body of evidences supporting a population of cold galaxies which are underrepresented in IRAS galaxies and hence  $70\ \mu\text{m}$  source counts (Stickel et al., 1998, 2000; Chapman et al., 2002; Patris et al., 2003; Dennefeld et al., 2005; Sajina et al., 2006; Amblard et al., 2010). However, based on available data there is no possible way to disentangle between the two mentioned possibilities (Le Borgne et al., 2009). Therefore, due to its simplicity, we try to find a modification in SED amplitudes which could improve the agreement between the "850-model" and observed source counts at SPIRE wavelengths and  $160\ \mu\text{m}$ , without affecting other wavelengths. In the following we first introduce the desired SED modification and then based on the redshift distribution of sources responsible for different source counts in our modified model, we try to constrain different properties a population of cold IR galaxies should possess to be equivalent to that SED modification.

Since the "850-model" is capable of fitting the observed  $70\ \mu\text{m}$  source counts, any modification of SED templates should be at wavelengths longer than  $70\ \mu\text{m}$  to leave this agreement intact. The  $850\ \mu\text{m}$  source counts that model produces should also remains the same which put an upper limit for the allowed wavelength range of any modification: since a significant fraction of observed counts at  $850\ \mu\text{m}$  is produced by sources at high redshifts, typically  $z > 2$  (see Figures 5.7 and 5.10), SED templates should not change at rest-frame

---

<sup>3</sup>Those works point out this issue only for the source counts at  $160\ \mu\text{m}$  since the observations at  $250\ \mu\text{m}$ ,  $350\ \mu\text{m}$  and  $500\ \mu\text{m}$  have been available only recently.

wavelengths longer than  $\sim 200 \mu\text{m}$ . Based on this ansatz, we searched for the amplitude and the wavelength range of SED changes which can bring the "850-model" intermediate wavelength counts close to the observed values. We found that if the SED templates we use in our model be amplified by a factor of 1.6 in the rest-frame range of  $70 < \lambda < 150 \mu\text{m}$ , the "850-model" can fit the SPIRE and  $160 \mu\text{m}$  source counts, without any change in already decent results at  $70 \mu\text{m}$ ,  $850 \mu\text{m}$  and  $1100 \mu\text{m}$ . Figure 5.10 illustrates the results produced by this modified "850-model" at different wavelengths (blue solid line) together with the results from Lagache et al. (2004) (orange dashed line) and Valiante et al. (2009) (gray dash-dotted line) models. The redshift distribution of sources responsible for those results are also shown in Figure 5.11.

The SED boost we implemented, recovers the observed  $160 \mu\text{m}$  counts by doubling the number of observable sources with  $S_{160} > 10\text{mJy}$ ; as illustrated in Figure 5.10, a significant fraction of those sources is distributed at redshifts  $z < 1$  which is also in agreement with recent observations (Jacobs et al., 2011; Berta et al., 2011). However, for the SPIRE source counts, the SED modification increases the number of observed sources with  $S_{\text{SPIRE}} < 100\text{mJy}$  which are mainly at redshifts  $0.5 < z < 2$ . Those redshift distributions together with the fact that the emission from cold dust with temperatures  $T \lesssim 30\text{K}$  peaks at  $\sim 100 \mu\text{m}$ , make our SED modification equivalent to adding a population of preferentially cold dusty galaxies which do not exist at high redshifts. The width of redshift range in which those objects can exist depends on their temperature range: a warmer population can have a wider redshift distribution but should have a typically higher redshifts to be invisible at  $70 \mu\text{m}$  while a colder population can only exist in low redshift ranges in order to not interfere with  $850 \mu\text{m}$  counts. Interestingly, this is in agreement with the required steep colour evolution implied by models constrained by  $160 \mu\text{m}$  and SPIRE band counts (see Table 5.2).

As one can see in Figure 5.10, some models like Lagache et al. (2004) and Valiante et al. (2009) not only produce enough observable sources in SPIRE and  $160 \mu\text{m}$  bands, but at some flux thresholds produce too many sources. However, we notice that both of those models need an additional mechanism to compensate for under-production of visible sources at intermediate wavelengths which mimic a population of cold sources only in low redshifts. For instance, Lagache et al. (2004) use a class of "cold galaxies" in their evolutionary model which are present only at low redshift,  $z < 0.5$ ; Lagache et al. (2003) show that this cold population is producing up to  $\sim 50\%$  of the observed  $170 \mu\text{m}$  flux which means without them the source count is reduced by the same factor, consistent with our finding and also other models (Le Borgne et al., 2009; Valiante et al., 2009). Similarly, Valiante et al. (2009) need to strongly modify the colour distribution of low redshift galaxies to correct for under-producing  $160 \mu\text{m}$  sources by a factor of  $\sim 5$ ; they modified the observed colour distribution of IRAS galaxies which they use as starting point (similar to our model, see §5.2.1) only for low redshift (i.e.  $z < 1$ ) objects, by assuming an asymmetric Gaussian distribution

which is 7 times broader on the "cold" side of distribution in comparison to the "warm" side (see §4.4 in Valiante et al. (2009)); even with this extreme modification their model does not completely match the observed  $160\ \mu\text{m}$  counts in addition to under-producing luminous  $70\ \mu\text{m}$  sources.

Finally, it is worth mentioning that our SED modification is not expected to change the best-fit models which are based on  $70\ \mu\text{m}$  and  $850\ \mu\text{m}$  since we used those source counts as constraints for our SED change search. However we double checked this issue by using the modified SED set and repeating the search in parameter space for a best-fit model which is constrained by source counts at  $70\ \mu\text{m}$ ,  $160\ \mu\text{m}$ , SPIRE bands and  $850\ \mu\text{m}$  and recovering the parameters which define the "850-model", for the best-fit model.

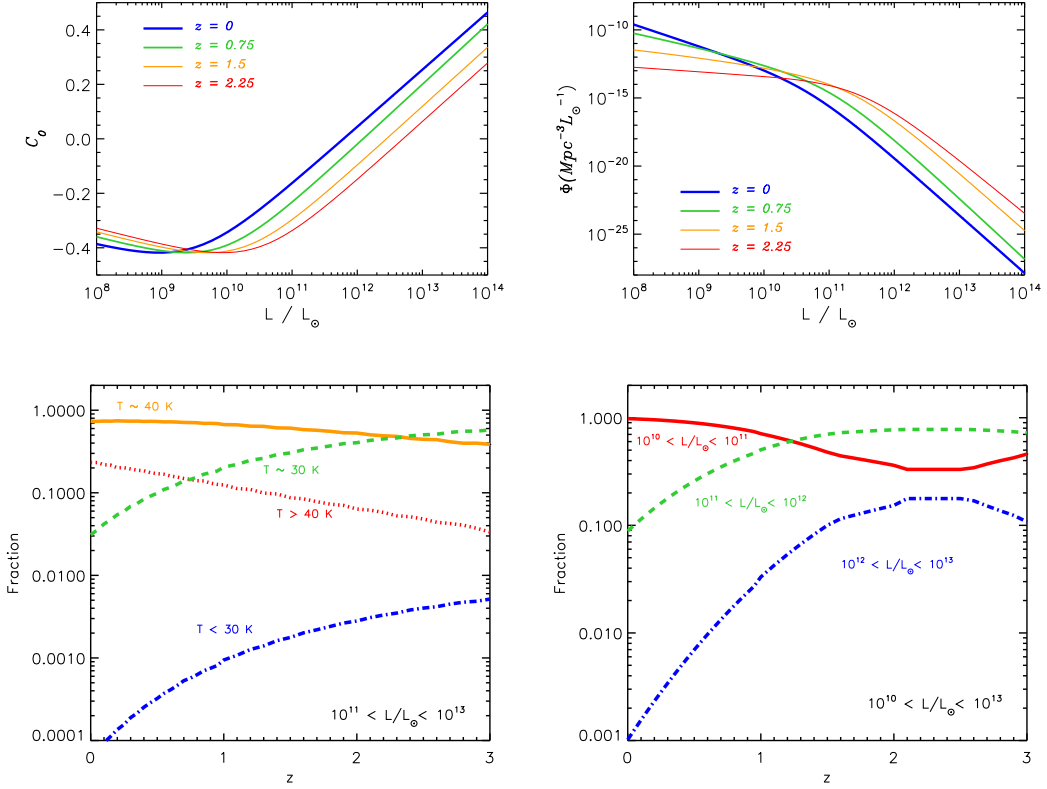
## 5.7 Discussion

In this section, we discuss different properties of our best-fit model. First we discuss different implications of our model for the evolving properties of the IR galaxies like their colour and luminosity distributions. Then, we compare our model with other existing models for IR and submm source counts.

### 5.7.1 The implied evolution scenario for dusty galaxies

Our best-fit model mimics the number density evolution of IR galaxies by employing a luminosity evolution together with changing slopes of LF with redshift. While the former changes the amplitude of LF, the latter acts to change the shape of LF properly to reproduce a number density history which matches the observed FIR and submm source counts. Although the good agreement between source counts which our best-fit model produces at different wavelengths and the observed numbers firmly supports our model, the implied results should be treated carefully, mainly due to the simplicity of the model: for instance, in our model we adopted the local LF of IR galaxies which is a simple dual power law function, and evolved its characteristic luminosity and slopes with redshift to distribute objects with different luminosities in redshift space correctly.

Although the functional forms we chose for evolution of those parameters are rather arbitrary, the actual distribution they produce at different redshifts is necessary to explain the properties of the observed sources. To emphasize this point, in Figure 5.12 we illustrate how the fractions of galaxies in different luminosity bins are evolving with redshift, together with the exact shape of the LF our model requires at different redshifts. As can be seen in the bottom right panel of Figure 5.12, at low redshifts fainter objects dominate the population of IR galaxies but at higher redshifts (i.e.,  $z > 2$ ), objects which are in brighter luminosity ranges take over and become more dominant. Specifically, this diagram shows that beyond  $z \sim 1$ , objects in the Luminous Infrared Galaxies (LIRGs) class (with IR luminosities between  $10^{11}$  and  $10^{12} L_{\odot}$ ) dominate the



**Figure 5.12:** Top: The evolution of CF (left) and LF (right) with redshift. The lines from thick to thin are respectively for redshifts  $z = 0$  (blue), 0.75 (green), 1.5 (orange) and 2.25 (red). Bottom: The change of the relative contribution of different populations of IR galaxies to their total density with redshift, as required to reproduce the observed source counts correctly. *left:* The fraction of objects with luminosities  $10^{11} \leq L/L_\odot \leq 10^{13}$  which have typical temperatures of  $T < 30$  K,  $\sim 30$  K,  $\sim 40$  K and  $T > 40$  K shown respectively with blue, green, orange and red lines. *right:* The fraction of total IR luminosity generated by objects with luminosities  $10^{10} \leq L/L_\odot \leq 10^{13}$ , which have luminosities of  $10^{10} \leq L/L_\odot < 10^{11}$ ,  $10^{11} \leq L/L_\odot < 10^{12}$  and  $10^{12} \leq L/L_\odot \leq 10^{13}$  shown respectively with red, green and blue lines.

obscured cosmic energy production (although in terms of numbers, fainter galaxies still dominate). The diagram also shows that UltraLuminous Infrared Galaxies (ULIRGs, with IR luminosities exceeding  $10^{12} L_{\odot}$ ), although increasing in importance towards high  $z$ , never dominate the cosmic energy production. These conclusions are in agreement with analyses of Spitzer data by Le Floch et al. (2005) for  $z < 1$  and by Magnelli et al. (2011) for  $z < 2.3$ .

Our best-fit model implies a change in the shape of the LF with redshift and we are the first to consider this possibility in a model of this type. This slope evolution implies that the faint end slope of the LF is flattening with increasing redshift (see the top right panel of Figure 5.12). At first sight it may seem surprising that our results are sensitive to the faint end slope at high redshifts, but the strong luminosity evolution combined with the strong negative  $K$ -correction brings sub- $L_*$  galaxies well within the region of detectability at  $850 \mu\text{m}$ . Nevertheless, this result depends strongly on the sub-mJy counts at  $850 \mu\text{m}$  which currently are derived from studies of gravitationally lensing clusters, and therefore sample limited cosmic volume. As such, these counts may be affected by cosmic variance and establishing their levels more firmly (e.g., with ALMA), is necessary for confirming our conclusion. However, we note the non-parametric estimates of the IR luminosity function at high redshifts (e.g., Chapman et al., 2005; Wardlow et al., 2011) do indicate flatter faint-end slopes than the local LF. These samples were selected at  $850 \mu\text{m}$  and may not be complete in IR luminosity, in particular they may be deficient in objects with high dust temperatures (e.g., Magdis et al., 2010). In order to settle this point, high- $z$  luminosity functions of IR luminosity-limited samples are required. Redshift surveys of Herschel-selected samples will be needed to construct such luminosity functions, and this will become feasible with facilities such as ALMA and CCAT. If confirmed, the flattening of the faint-end slope of the LF towards higher redshifts requires a physical explanation, which perhaps could be found using physically motivated models and simulations. One possibility is that a higher metagalactic ultraviolet flux at higher redshifts would suppress the development of a star-forming interstellar medium in low-mass galaxies. Another possibility is a stellar-mass-dependent evolution in  $M_{\text{dust}}/M_{\text{stars}}$  towards higher redshifts. Such an evolution could result from the decreasing overall metallicity towards high  $z$  combined with the net effect over time of the buildup of dust as a result of stellar evolution and its consumption by star formation. The latter model can be tested observationally using a combination of Herschel data and multi-band optical imaging (Bourne et al., in prep.).

We also showed that our model requires a colour evolution to reproduce the observed source counts. This colour evolution implies that objects with the same luminosities have lower typical dust temperatures at higher redshifts. In the bottom left panel of Figure 5.12 we showed how the fraction of objects with different temperatures is evolving with redshift for all the objects which have intrinsic luminosities between  $10^{11} < \frac{L}{L_{\odot}} < 10^{13}$ . While at low redshifts a

population of objects which have typical temperatures<sup>4</sup> of 40K, dominates the colour distribution, at higher redshifts the cooler typical temperature of 30K is dominating. Moreover, while at low redshifts objects with warm dust temperatures are more common than the very cold objects, the situation is the opposite at earlier times. Although it is difficult to observe the evolution of dust towards cooler temperatures because of different selection effects, there is observational evidence for the implied colour (i.e., temperature) evolution (Chapman et al., 2005; Pope et al., 2006; Chapin et al., 2009; Symeonidis et al., 2009; Seymour et al., 2010; Hwang et al., 2010; Amblard et al., 2010).

The luminosity and colour evolutions discussed above, show up in redshift distributions of source counts at different wavelengths. Figure 5.11 illustrates the buildup of source counts at different wavelengths by contribution from different redshift bins. It is evident that the longer wavelengths are reflecting the distribution of higher redshift dusty galaxies while shorter wavelengths depend more on IR galaxies at lower redshifts. For instance 850  $\mu\text{m}$  number counts mainly consist of galaxies with redshifts higher than  $z \sim 2$ , especially at fluxes  $\sim 10 - 20\text{mJy}$ ; in addition, as shown in the bottom right panel of Figure 5.12, at redshifts  $z \sim 2$  the luminosity distribution of IR galaxies is not evolving considerably and even at higher redshifts the fraction of fainter sources, which have lower dust temperatures, goes up again in comparison to the brightest objects. Moreover, for two objects with the same redshifts and intrinsic IR luminosities, the colder one will be brighter at observed 850  $\mu\text{m}$ . This means the visible 850  $\mu\text{m}$  sources at higher flux thresholds are preferentially colder than sources with fainter observed fluxes which are distributed typically at lower redshifts (see the increasing fraction of objects in  $1 < z < 2$  redshift bin going towards lower fluxes in the top right panel of Figure 5.11). This is also in agreement with our experiments which show that at the bright end, 850  $\mu\text{m}$  number counts are very sensitive to the slope of the colour evolution in our model, which is in fact the dominant mechanism to produce those counts. Moreover, this implied broader temperature distribution of observed sources at fainter flux thresholds and shorter wavelengths is in agreement with observations which show bright 850  $\mu\text{m}$  population (i.e.  $> 4\text{mJy}$ ) is highly biased towards cold dust temperatures while fainter sources (i.e.  $1 - 4\text{mJy}$ ) also contain an increasing fraction of more luminous objects with lower redshifts but warmer dust temperatures which makes them the main contributors to the 250  $\mu\text{m}$  source counts (Chapman et al., 2004, 2010; Magnelli et al., 2010; Casey et al., 2009, 2011). Consequently, at  $z \sim 2$  the density of farIR selected ULIRGs is approximately 2 times higher than that of 850  $\mu\text{m}$ -selected ULIRGs. This is consistent with our model: a typical ULIRG at redshift  $z \sim 2$  will be visible at 850  $\mu\text{m}$  with a flux brighter than a few mJys and at 250  $\mu\text{m}$  brighter than  $\sim 100\text{mJy}$ ; on the other hand, roughly  $\sim 20 - 30\%$  of objects with few mJy fluxes at 850  $\mu\text{m}$  are within redshifts  $z \sim 1 - 2$  while around  $\sim 50\%$  of objects with  $\sim 100\text{mJy}$  fluxes at 250  $\mu\text{m}$  are in the same redshift bin

---

<sup>4</sup>We assign a single temperature to each SED based on its colour and finding a modified black body radiation with a fixed emissivity,  $\beta = 1.5$ , which can produce that colour.



(see Figure 5.11).

While the luminosity and colour distribution our best-fit model requires is consistent with observed  $70\ \mu\text{m}$ ,  $850\ \mu\text{m}$  and  $1100\ \mu\text{m}$  source counts, they are not sufficient to produce enough sources at observed wavelengths in between. The inconsistency between models which successfully produce  $70\ \mu\text{m}$  and  $850\ \mu\text{m}$  counts and their results at  $160\ \mu\text{m}$ , is also reported in other works and has been corrected by including a population of cold galaxies at low redshifts (Lagache et al., 2003, 2004; Le Borgne et al., 2009; Valiante et al., 2009). While it is important to note the under-production of  $160\ \mu\text{m}$  counts in models, can be corrected equally by a modifying SED templates instead of introducing a new population (see also Le Borgne et al. (2009)), there is some observational evidence for the existence of a cold population at low redshifts which is under-represented in IRAS sample and is often associated with bright spiral galaxies (Stickel et al., 1998, 2000; Chapman et al., 2002; Patris et al., 2003; Dennefeld et al., 2005; Sajina et al., 2006; Amblard et al., 2010).

Recently, flux density measurements at  $250$ ,  $350$  and  $500\ \mu\text{m}$  have become available for large samples of local galaxies from surveys with SPIRE on the Herschel Space Observatory (Eales et al., 2010; Oliver et al., 2010; Clements et al., 2010). In addition to problems at  $160\ \mu\text{m}$ , we also noticed the inconsistency between our best-fit model and the source counts provided by SPIRE observations. However, we resolved this issue by modifying our SED templates to be able to reproduce the source count simultaneously at  $70\ \mu\text{m}$ ,  $160\ \mu\text{m}$ , SPIRE band,  $850\ \mu\text{m}$  and  $1100\ \mu\text{m}$ . There is also observational evidence for a population of cold galaxies residing at low redshifts (equivalent to modified SEDs) in Herschel-selected samples: using a subsample with spectroscopic or reliable photometric redshifts from Herschel ATLAS survey, Amblard et al. (2010) performed isothermal graybody fits to low-redshift galaxies detected from  $70$  to  $500\ \mu\text{m}$ , resulting in an IR luminosity-temperature relation offset to significantly lower temperatures when compared to the IRAS-based relation derived by Chapman et al. (2003). The new relation found by Amblard et al. (2010) is consistent with earlier work by Dye et al. (2009) based on BLAST data. It is also important to realize that these results do not imply that the IRAS-based dust temperature fits are incorrect (in fact, they are often supplemented with measurements at longer wavelengths) but they imply that an IRAS-based selection is biased towards warmer objects.

## 5.7.2 Our best-fit model and previous models

There are several phenomenological models in the literature which try to reproduce the properties of IR galaxies at different wavelengths, with different levels of complexity (e.g., Blain & Longair (1993); Guiderdoni et al. (1997); Blain et al. (1999); Chary & Elbaz (2001); Rowan-Robinson (2001); Dole et al. (2003); Lagache et al. (2004); Lewis et al. (2005); Le Borgne et al. (2009); Valiante et al. (2009); Bethermin et al. (2011)). In general, those models

use an assumed form of luminosity evolution together with a density evolution to mimic the evolution of IR galaxy distributions. However, we have limited ourselves to a pure luminosity evolution with no density evolution, and as pointed out in §5.2.2, the amount of density evolution allowed by the integrated CIB is small but does not have to be zero. On the other hand, we use evolving faint and bright ends slopes in our luminosity function. At first sight this may look like a simple substitution of free parameters; we tried to investigate this by trying to substitute the slope evolution of our model with a density evolution. However, the resulting fit with density evolution and in absence of slope evolution is significantly inferior to the fit obtained with pure luminosity evolution.

Another usual practice in modeling the infrared and submm source counts is to use only one or a few SED templates to represent the whole galaxy population. This approach neglects the observed colour distribution of local IR galaxies and does not leave any possibility for colour evolution. However, similar to Valiante et al. (2009), in our model we use a complete set of SED templates which are shown to be representative of local IR galaxies. This choice enabled our model to explore the evolution in colours of IR galaxies in addition to their luminosities.

Our model also differs from other works in the literature in calculating the source counts for a given evolutionary scenario: we calculate the source count for a given model by computing the probability of observing different sources for different flux thresholds. The direct consequence of this new approach is a fast calculation routine which enables us to calculate the source counts at very bright observed fluxes, where Monte-Carlo based methods are very inefficient due to the rarity of such objects, and therefore sometimes end up with very noisy results (see the bright source counts produced by Valiante et al. (2009) in Figure 5.10). Moreover, our fast algorithm is an important advantage when it comes to searching the parameter space for the best-fit model.

A comparison between our best-fit model and Lagache et al. (2004), as a model without colour distribution and evolution and Valiante et al. (2009) as a model with colour distributions is shown in Figure 5.8 and Figure 5.9. While at  $1100\ \mu\text{m}$ , Valiante et al. (2009) produces  $\sim 2$  times more visible objects than what our model produces, all models do reasonably well in accounting for the cumulative  $850\ \mu\text{m}$  number counts. The results from Valiante et al. (2009) and our model are very close at SPIRE wavelengths but the Lagache et al. (2004) model overpredicts the bright counts at  $250\ \mu\text{m}$ . At  $160\ \mu\text{m}$ , the Valiante et al. (2009) model over-produces the faint counts and under-produces the bright objects. However, Lagache et al. (2004) fit the data better, while slightly over-produces the counts for intermediate to bright flux thresholds. Finally, at  $70\ \mu\text{m}$ , where all the models are expected to fit the data since they use it as a starting point, the Valiante et al. (2009) source counts deviate from observations by over-producing the faint counts in expense of producing too few bright objects (probably because of a too extreme modification in colour distributions which is required in their model to compensate for a factor of  $\sim 5$  under-production of  $160\ \mu\text{m}$



sources).

## 5.8 Conclusions

We have described a backward evolution model for the IR galaxy population, with a small number of free parameters, emphasizing which parameters are constrained by which observations. We also introduced a new algorithm for calculating source counts for a given evolutionary model by direct integration of probability distributions which is faster than using Monte-Carlo sampling. This is an important advantage for searching large volumes of parameter space for the best-fit model.

While most of the earlier works used only one or a handful of SED templates to represent the whole population of IR objects, we used a library of IR SEDs which are able to match the IR properties of the large variety of observed star-forming objects. This approach is necessary in order to model the colour evolution of IR galaxies in addition to produce simultaneously the counts and the redshift distributions at wavelengths shorter than  $850\ \mu\text{m}$ .

Contrary to some other models, we assumed a negligible contribution from AGN in our SED templates, noting the inclusion of AGN is only necessary for reproducing the properties of IR galaxies at very short IR wavelengths<sup>5</sup> which could also be sensitive to other modeling difficulties such as the PAH contribution to the SEDs.

We used available  $850\ \mu\text{m}$  source counts together with the redshift distribution of submm galaxies to constrain our best-fit model. At  $850\ \mu\text{m}$ , due to the K-correction, the source count is sensitive to the evolution of IR galaxies in a wide redshift range and out to very high redshifts. We showed that there is a degeneracy between the rate by which the characteristic luminosity of IR galaxies should increase to reproduce the source count and the maximum redshift out to which this increase should be continued; we resolve this degeneracy by requiring the model to reproduce the observed redshift distribution of submm galaxies. Moreover, we showed that our model requires a colour evolution towards cooler typical dust temperatures at higher redshifts. The employed colour evolution is similar to that used by Valiante et al. (2009) however, our best-fit model predicts a somewhat stronger colour evolution than that proposed by these authors.

Another important feature of our model is that the best-fit is obtained using pure luminosity evolution but mildly evolving high-luminosity and low-luminosity slopes in the LF. Since high-luminosity sources are rare, the evolution in the high-luminosity slope is of little consequence. However, the evolution of the low-luminosity slope affects large numbers of galaxies and if confirmed, this effect must have a physical origin, which can be addressed using numerical

---

<sup>5</sup>For instance at  $24\ \mu\text{m}$  where our model under-produces the counts at flux thresholds  $0.1 < S_{24} < 10\text{mJy}$  by a factor of  $\sim 1.5 - 2$  but matches the observed data at fainter and brighter fluxes

simulations of the evolution of the galaxy population, as well as with a combination of deep Herschel and optical imaging.

The 850  $\mu\text{m}$ -constrained best-fit model is consistent with observed 1100  $\mu\text{m}$  and 70  $\mu\text{m}$  source counts, which confirm the consistency of the implied colour-luminosity-redshift distribution at both low and high redshifts. However, this model under-produces the observed source counts at intermediate wavelengths, namely at 160  $\mu\text{m}$  and SPIRE bands. To resolve this issue, we used the observed data at different wavelengths to find best-fit models which can reproduce their observed source counts. While the best-fit models constrained at 70  $\mu\text{m}$  and 850  $\mu\text{m}$  are consistent with each other and also 1100  $\mu\text{m}$ , the implied evolutions for models capable of reproducing observed counts at other wavelengths are too diverse to be reconciled in a single model; specifically they need too strong colour evolutions which contradicts 850  $\mu\text{m}$  observations. While the inconsistency at 160  $\mu\text{m}$  has been reported in earlier works (Le Borgne et al., 2009; Valiante et al., 2009), we are the first to report it for 250  $\mu\text{m}$ , 350  $\mu\text{m}$  and 500  $\mu\text{m}$ . We showed that the source counts at these wavelengths can be reproduced consistently, by adopting the best-fit model which produces correct 70  $\mu\text{m}$ , 850  $\mu\text{m}$  and 1100  $\mu\text{m}$  source counts, together with a modification in SED templates which is equivalent to the existence of a cold population of dusty galaxies at low to intermediate redshifts which are under-represented in IRAS data. Besides the fact that there is some observational evidence for the existence of such galaxies (Stickel et al., 1998, 2000; Chapman et al., 2002; Patris et al., 2003; Dennefeld et al., 2005; Sajina et al., 2006; Amblard et al., 2010), this assumption is similar to what other models had to assume in order to reproduce adequate 160  $\mu\text{m}$  sources (Lagache et al., 2003, 2004; Valiante et al., 2009).

It is important to keep in mind that phenomenological models like what we described in this chapter, are mainly simple mathematical forms which relate different observations consistently rather than being physical models with explanatory power. However, their performance at different wavelengths and the distribution of sources they require for different redshifts can be used as their main predictions which also could be used to test their validity. While we used the redshift distribution of submm galaxies to constrain our model, we note that the observed redshift distribution of other wavelengths, if available, are in agreement with our best-fit model predictions (Jacobs et al., 2011; Berta et al., 2011).

Additional information including tabulated data for differential and cumulative source counts at different wavelengths and their redshift distributions is available at <http://www.strw.leidenuniv.nl/genesis/>

## Acknowledgments

We thank the anonymous referee for valuable comments which improved the original version of the paper this chapter is based on. We thank D. Dale, H. Rottgering, J. Schaye and M. Shirazi for useful discussions. During the early

---

stages of this work, AR was supported by a Huygens Fellowship awarded by the Dutch Ministry of Culture, Education and Science.

## References

- Alexander D.M., et al., 2005, *ApJ*, 632, 736  
Amblard A., et al., 2010, *A&A*, 518, L9  
Austermann J.E., et al., 2009, *MNRAS*, 393, 1573  
Austermann J.E., et al., 2010, *MNRAS*, 401, 160  
Barger A.J., et al., 1999, *ApJ*, 518, L5  
Berta B., et al., 2011, *A&A*, arXiv:1106.3070  
Bertoldi F., et al., 2007, *ApJS*, 172, 132  
Bethérmin M., et al., 2010, *A&A*, 512, 78  
Bethérmin M., et al., 2011, *A&A*, 529, 4  
Blain A.W., Longair M.S., 1993, *MNRAS*, 264, 509  
Blain A.W., et al., 1999, *MNRAS*, 302, 632  
Borys, C., et al., 2003, *MNRAS*, 344, 385  
Casey C.M., et al., 2009, *MNRAS*, 399, 121  
Chapin E.L., et al., 2009, *MNRAS*, 393, 653  
Chapman S.C., et al., 2002, *MNRAS*, 335, 17  
Chapman S.C., et al., 2003, *ApJ*, 588, 186  
Chapman S.C., et al., 2004, *ApJ*, 614, 671  
Chapman S.C., et al., 2005, *ApJ*, 622, 772  
Chapman S.C., et al., 2010, *MNRAS*, 409, 13  
Casey C.M., et al., 2009, *MNRAS*, 399, 121  
Casey C.M., et al., 2011, *MNRAS*.tmp.884C  
Chary R., Elbaz D., 2001, *ApJ*, 556, 562  
Clements D.L., et al., 2010, *A&A*, 518, 8  
Coppin K., et al., 2006, *MNRAS*, 372, 1621  
Cowie K., et al., 2006, *MNRAS*, 372, 1621  
Dale D., et al., 2001, *ApJ*, 549, 215  
Dale D., Helou G., 2002, *ApJ*, 576, 159  
Dennefeld M., et al., 2005, *A&A*, 440, 5  
Dole H., et al., 2003, *ApJ*, 585, 617  
Dye S., et al., 2009, *ApJ*, 703, 285  
Eales S., et al., 2010, *PASP*, 122, 499  
Efstathiou A., et al., 1995, *MNRAS*, 277, 1134  
Fadda D., et al., 2010, *ApJ*, 719, 425  
Fixsen D. J., et al., 1998, *ApJ*, 508, 123  
Gispert R., et al., 2000, *A&A*, 360, 1  
Greve, T. R., et al., 2004, *MNRAS*, 354, 779  
Guiderdoni B., et al., 1997, *Nature*, 390, 257  
Hatsukade B., et al., 2011, *MNRAS*, 411, 102

- Hopkins A.M., Beacon J.F., 2006, *ApJ*, 651, 142  
Hwang H.S., et al., 2010, *MNRAS*, 409, 75  
Jacobs B.A., et al., 2011, *AJ*, 141, 110  
Jauzac J.M., et al., 2010, *A&A*, sub.  
Johansson D., et al., 2011, *A&A*, 527, A117  
Khan S.A., et al., 2007, *ApJ*, 665, 973  
Knudsen K.K., et al., 2008, *MNRAS*, 384, 1611  
Lagache G., et al., 2003, *MNRAS*, 338, 555  
Lagache G., et al., 2004, *ApJS*, 154, 112  
Lagache G., et al., 2005, *ARA&A*, 43, 727  
Le Borgne D., et al., 2009, *A&A*, 504, 727  
Le Flo'ch E., et al., 2005, *ApJ*, 632, 169  
Lewis G.F., et al., 2005, *ApJ*, 621, 32  
Lutz D., et al., 2005, *ApJ*, 632, 13  
Magdis, G.E., et al., 2010, *MNRAS*, 409, 22  
Magnelli, B., et al., 2010, *A&A*, 518, 28  
Magnelli, B., et al., 2011, *A&A*, 528, 35  
Menendez-Delmestre K., et al., 2009, *ApJ*, 699, 667  
Mullaney J.R., et al., 2011, *MNRAS*, arXive:1106.4284  
Oliver S.J., et al., 2010, *A&A*, 518, 21  
Patanchon G., et al., 2009, *ApJ*, 707, 1750  
Patris J., et al., 2003, *A&A*, 412, 349  
Pope A., et al., 2006, *MNRAS*, 370, 1185  
Pope A., et al., 2008, *ApJ*, 675, 1171  
Puget J.L., et al., 1996, *A&A*, 308,L5  
Rowan-Robinson M. 2001, *ApJ*, 549, 745  
Saunders W., et al., 1990, *MNRAS*, 242, 318  
Sajina A., et al., 2006, *MNRAS*, 369, 939  
Schaye J., et al., 2010, *MNRAS*, 402, 1536  
Scott K.S., et al., 2010, *MNRAS*, 405, 2260  
Seymour N., et al., 2010, *MNRAS*, 402, 2666  
Smail I., et al., 1997, *ApJ*, 490, L5  
Smail I., et al., 2002, *MNRAS*, 331, 495  
Stickel M., et al., 1998, *A&A*, 336, 116  
Stickel M., et al., 2000, *A&A*, 359, 865  
Swinbank A.M., et al., 2004, *ApJ*, 617, 64  
Symeonidis M., et al., 2009, *MNRAS*, 397, 1728  
Valiante E., et al., 2007, *ApJ*, 660, 1060  
Valiante E., et al., 2009, *ApJ*, 701, 1814  
Vieira, J. D., et al., 2010, *ApJ*, 719, 763  
Wardlow, J. L., et al., 2011, *MNRAS*, tmp.917w  
Webb T.M., et al., 2003, *ApJ*, 587, 41  
Weiss A., et al., 2009, *ApJ*, 707, 1201

## Appendix A: Some numerical details

For each specific evolution model, the source count at a given flux threshold and wavelength can be calculated based on equation (5.13), where the integration should be performed over all possible luminosities, colours and redshifts. As mentioned in §5.2.4, we do this by splitting possible colour, luminosity and redshift ranges into very small bins, assuming that in each bin the related variable is not changing significantly.

The finite number of SED models we are using automatically splits the colour range into 64 bins between  $0.29 \leq R(60,100) \leq 1.64$  (see §5.2.3). We also use logarithmically spaced bins to split the possible luminosity range of  $10^9 L_\odot \leq L \leq 10^{14} L_\odot$  into 100 bins in our calculation. This logarithmic scale which makes the integration roughly insensitive to the number of luminosity bins, is chosen to cope with the exponential nature of luminosity function where faint objects are much more numerous than luminous ones. It is also worth mentioning that the source count calculation is not sensitive to the minimum or maximum luminosity which is used in integration, if the used luminosity range covers the important  $10^{10} - 10^{13} L_\odot$  luminosity range; for instance, using  $L_{\min} = 10^7 L_\odot$  instead of  $L_{\min} = 10^9 L_\odot$  as the minimum possible luminosity, does not change any of the source count calculations we are presenting in this paper.

As discussed in §5.2.4, the uniform distribution of galaxies in redshift space is assured by the algorithm we are using, independent of the size of redshift bins. But, for the precise calculation of the K-correction and evolution functions, we split the redshift range of  $0 \leq z \leq 8$  using bin sizes equal to  $\Delta z = 0.01$ . However we noted it is possible to use even bigger redshift bins (e.g.  $\Delta z = 0.1$ ) without any significant change in the results.



# NEDERLANDSE SAMENVATTING

De theorie die de vorming van sterrenstelsels beschrijft, omvat natuurkunde waarin zowel de grootste als de kleinste schalen in het heelal een belangrijke rol spelen. Het enorme dynamische bereik van de verschillende fysische grootheden (zoals lengte, massa en leeftijd) die een sterrenstelsel karakteriseren, maakt het modelleren van deze complexe systemen tot een grote uitdaging. Niettemin zou de vorming en evolutie van sterrenstelsels in principe te begrijpen moeten zijn vanuit de bestaande fysische wetten. Het nauwkeurig beschrijven en verklaren van de waargenomen trends vanuit deze wetten is tot op heden één van de belangrijkste doelen van het onderzoeksgebied. Dankzij het werk van vele grote wetenschappers hebben we nu een tamelijk goed beeld van hoe sterrenstelsels vormen en zich ontwikkelen en kunnen we met een redelijke nauwkeurigheid verklaren wat we zien van deze systemen in het heelal. Toch zijn er nog tal van fenomenen die we niet doorgronden met de huidige theorie van sterrenstelsels, omdat, evenals in andere gebieden van de natuurwetenschap, iedere stap vooruit in ons begrip van de vorming en evolutie van sterrenstelsels onherroepelijk leidt tot nieuwe ononttrafelde problemen en uitdagingen.

## Neutraal waterstof in galactische ecosystemen

Sterrenstelsels worden beïnvloed door aan de ene kant de zwaartekracht, die verantwoordelijk is voor de vorming van halo's, het aantrekken van nieuw materiaal (voornamelijk gas) en het bijhouden van de 'baryonische' en donkere materie waauit een stelsel is opgebouwd, en aan de andere kant de verschillende 'feedback' mechanismen die juist tegen de zwaartekracht in werken en proberen de opgebouwde structuren te ontbinden. De interactie tussen deze strijdige mechanismen creëert complexe ecosystemen in en rondom sterrenstelsels en drukt een stempel op de verdeling van baryonen in de nabije omgeving van deze stelsels (het zogenaamde 'circumgalactische medium', CGM). De verdeling van baryonen, en de verdeling van neutraal waterstof (HI) in het bijzonder, verschaft ons waardevolle informatie over de ontstaansgeschiedenis van deze sterrenstelsels. De reden dat HI een belangrijke rol speelt in het onderzoek, is dat HI de voornaamste brandstof vormt voor de vorming van moleculaire gaswolken, de geboorteplaatsen van sterren. Derhalve is onderzoek naar de verdeling van HI en hoe deze verandert met de tijd cruciaal voor ons begrip van de verschillende aspecten van stervorming.

We weten dat vlak na de Big Bang het, in eerste instantie hete, plasma begint af te koelen naarmate het heelal uitdijt en dat protonen en elektronen uiteindelijk recombineren tot neutraal waterstof. Op  $z \sim 6$  raakt waterstof, het meest voorkomende element in het heelal, echter opnieuw in hoge mate geïoniseerd ('reïonisatie') door de vorming van de eerste generatie sterren en sterrenstelsels. De typische afstand die fotonen kunnen afleggen voordat ze worden geabsorbeerd, neemt vanaf dat moment aanzienlijk toe naarmate de tijd vordert, omdat aan de ene kant het heelal uitdijt en dus minder dicht wordt en aan de andere

kant de activiteit van stervorming in het heelal toeneemt. De vele bronnen van ioniserende straling, die op grote schaal relatief uniform verdeeld zijn, zorgen voor een min of meer uniforme achtergrond van ultraviolette straling (afgekort als 'UVB'), die op  $z \geq 3$  wordt gedomineerd door fotonen afkomstig van sterren en op lage roodverschuiving door het licht van quasars. Als gevolg van deze alom aanwezige bron van straling zijn de meeste waterstofatomen blijvend geïoniseerd vanaf het moment van reïonisatie op  $z \sim 6$ .

Hoe complex de verschillende processen zijn die de ionisatietoestand van waterstof bepalen, hangt sterk af van de kolomdichtheid van het waterstof. Bij lage dichtheden (d.w.z.  $N_{\text{HI}} \leq 10^{17} \text{ cm}^{-2}$ , wat overeenkomt met het zogenaamde 'Lyman- $\alpha$  forest') is HI in hoge mate geïoniseerd door de UVB-straling en overwegend transparant voor ioniserende straling van de UVB dan wel van andere bronnen. In dit geval kunnen we het systeem beschouwen als 'optisch dun' en kunnen we werken in de 'optisch dunne limiet'. Bij hogere dichtheden (d.w.z.  $N_{\text{HI}} \geq 10^{17} \text{ cm}^{-2}$ , wat overeenkomt met de zogenaamde 'Lyman Limit' en 'Damped Lyman- $\alpha$ ' systemen) is het gas echter optisch dik en daarom overwegend neutraal. De ionisatietoestand van HI in dergelijke systemen is sterker afhankelijk van de mate waarin de verschillende aspecten, zoals de schaduwgevende werking van optisch dik materiaal en de fluctuaties van de UVB-straling op kleine schaal, een rol spelen bij het stralingstransport.

Het onderzoeken van de HI-verdeling in het heelal blijkt geen gemakkelijke opgave. In het lokale heelal kan de hoeveelheid HI in sterrenstelsels worden gemeten door gebruik te maken van 21 cm-emissie, maar op hogere roodverschuiving is dit vooralsnog niet mogelijk. Het wachten is op de komst van krachtigere telescopen, zoals het *Square Kilometer Array*. Op  $z \leq 6$ , dus na reïonisatie, is het mogelijk HI te bestuderen met behulp van 21 cm-absorptie. We kunnen de verdeling van het gas bepalen door spectra van heldere achtergrondbronnen zoals quasars te analyseren en te kijken naar de absorptiekenmerken van HI-systemen door zich tussen ons en de bron bevinden. Op hoge roodverschuiving is het bestuderen van deze absorptiekenmerken daarom een goed alternatief voor metingen in emissie. De enorme afstanden tussen de meeste absorberende objecten en hun achtergrondbronnen maken het onwaarschijnlijk dat ze op enig fysische manier gerelateerd zijn, iets wat de absorptielijnen in een spectrum zou kunnen beïnvloeden. Dit geeft ons de mogelijkheid om de verdeling van HI langs de gezichtslijn nauwkeurig te reconstrueren.

De laatste tientallen jaren hebben verscheidene observationele onderzoeken zich toegespitst op het bepalen van de statistische eigenschappen van de HI-verdeling in het heelal. Dankzij een significante toename van het aantal quasars dat is waargenomen en een verbetering van de observatietechnieken spannen deze waarnemingen nu een groot bereik in kolomdichtheid en roodverschuiving. Het is daarom belangrijk om de eigenschappen van de HI-verdeling ook in kosmologische simulaties te bestuderen, die ons een beter begrip van de waargenomen trends zullen geven en ons bovendien zullen helpen de verschillende observaties in de context van de huidige theorie van sterrenstelsels te plaatsen.



---

Hier zal ik mij op richten in dit Proefschrift. Als onze simulaties overeenkomen met de observaties, kunnen we deze gebruiken om te voorspellen hoe toekomstige waarnemingen eruit zullen zien. Deze voorspellingen kunnen dan gebruikt worden aan de ene kant om de juistheid van de onderliggende modellen in de simulaties te testen, en waar nodig verbeteringen aan te brengen, en aan de andere kant om te onderzoeken welke rol de verschillende fysische processen spelen in de vorming en evolutie van sterrenstelsels.

## Dit Proefschrift

Hedendaagse observationele onderzoeken gaan verder dan het bestuderen van de sterren in sterrenstelsels alleen, maar richten zich ook op de verschillende belangrijke, maar zeer complexe, processen die de verdeling van gas in en rondom stelsels beïnvloeden. Ook de hydrodynamische kosmologische simulaties die gebruikt worden om de vorming van sterrenstelsels te onderzoeken, verbeteren aanzienlijk door het gebruik van steeds betere numerieke technieken, sterker onderbouwde fysische modellen en hogere resoluties. Het is belangrijk, zo niet essentieel, om observaties en simulaties met elkaar te blijven vergelijken om aan de ene kant een beter begrip te krijgen van de observationele resultaten en aan de andere kant de simulaties te testen en te verbeteren.

De verdeling van HI en hoe deze verandert met de tijd is nauw verbonden met de verschillende aspecten van stervorming. Derhalve is het begrip en modelleren van de HI-verdeling van cruciaal belang bij het onderzoek naar de vorming en evolutie van sterrenstelsels. In dit Proefschrift zal ik mij voornamelijk richten op het bestuderen van de kosmische verdeling van neutraal waterstof met behulp van hydrodynamische kosmologische simulaties. We zullen dit doen door simulaties die gebaseerd zijn op de *OverWhelmingly Large Simulations*, te combineren met nauwkeurige beschrijvingen van stralingstransport, waarbij we rekening houden met de verschillende foto-ionisatieprocessen.

In **Hoofdstuk 2** beschrijven we de effecten van de metagalactische UVB-straling en diffuse recombinatiestraling op het stralingstransport en de ionisatie van waterstof op  $z = 5 - 0$ . We zullen ons richten op kolomdichtheden van  $N_{\text{HI}} > 10^{16} \text{ cm}^{-2}$ , omdat hierbij de effecten van stralingstransport vooral van belang zijn. Door ruim 12 miljard jaar van de evolutie van de HI-verdeling te modelleren tonen we aan dat de voorspelde verdeling van kolomdichtheden uitstekend overeenkomt met de waargenomen verdeling en slechts een matige verandering vertoont van  $z = 5$  tot  $z = 0$ . We vinden ook dat op  $z \geq 1$  de UVB de belangrijkste bron van ioniserende straling is, maar dat op lagere roodverhuiving botsingsionisatie een steeds grotere rol begint te spelen, wat gevolgen heeft voor de schaduwwerking van optisch dik materiaal. Op basis van onze simulaties presenteren we een aantal afgeleide verbanden die gebruikt kunnen worden om heel precies de fractie neutraal waterstof te kunnen berekenen zonder rekening te houden met stralingstransport. Gezien de complexiteit van de

stralingstransportsimulaties zullen deze verbanden bijzonder nuttig zijn voor de volgende generatie van kosmologische simulaties op hoge resolutie.

Stervorming vindt typisch plaats in omgevingen van zeer hoge kolomdichtheid, waar het gas optisch dik en daarom afgeschermd van externe ioniserende straling is. De lokale sterren vormen daarom een belangrijke bron van ioniserende straling bij deze dichtheden. Het simuleren van deze effecten is echter geen gemakkelijke opgave doordat het aantal stralingsbronnen over het gehele simulatievolume extreem groot is. Verschillende onderzoeken vinden tot op heden dan ook verschillende, vaak geenszins overtuigende, resultaten wat betreft de invloed van straling afkomstig van sterren op de HI-verdeling. In **Hoofdstuk 3** bieden we een oplossing voor dit probleem door kosmologische simulaties te combineren met stralingstransport door gebruik te maken van TRAPHIC, een code die ontwikkeld is om op een efficiënte manier om te gaan met grote aantallen stralingsbronnen. We simuleren de ioniserende straling van zowel sterren als van de UVB en recombinatieprocessen. We tonen aan dat de straling van lokale sterren significant van invloed kan zijn bij kolomdichtheden zoals die voorkomen in 'Damped Lyman- $\alpha$ ' (DLA) en 'Lyman Limit' (LL) systemen. We laten bovendien zien dat de voornaamste reden dat de resultaten van eerdere werken niet altijd overeenkwamen, het gebrek aan een voldoende hoge resolutie is. Dit probleem lossen wij echter op door stervormende 'deeltjes' in de simulatie als ionisatiebronnen te gebruiken. We tonen ook aan dat wanneer het interstellair medium (ISM) in kosmologische simulaties niet volledig is opgelost, dit het modelleren van de eigenschappen van sterke DLA's (d.w.z.  $N_{\text{HI}} \geq 10^{21} \text{ cm}^{-2}$ ) in de weg staat.

Objecten die zorgen voor sterke HI-absorptielijnen als DLA's zijn hoogst waarschijnlijk representatief voor het koude gas dat zich in, of in de buurt van, het ISM van sterrenstelsels op hoge roodverschuiving bevindt. Deze objecten maken het daarom mogelijk een verzameling sterrenstelsels te selecteren op basis van hun absorptiekenmerken en aan de hand hiervan het ISM te bestuderen, met name hoe dit eruit zag in de vroege stadia van de vorming en evolutie van sterrenstelsels. Omdat observationele onderzoeken doorgaans beperkt worden door het kleine aantal sterk absorberende bronnen en bovendien voor de lage-massa stelsels detecties missen in emissie, is het onderzoeken van de relatie tussen de absorberende objecten en de sterrenstelsels waarin zij zich bevinden, echter verre van gemakkelijk en is het noodzakelijk om gebruik te maken van kosmologische simulaties om de relatie tussen deze twee te doorgronden. In **Hoofdstuk 4** gebruiken we hydrodynamische kosmologische simulaties, waarvan we in Hoofdstuk 2 hebben laten zien dat zij goed overeenkomen met wat we zien in observaties van de HI-verdeling, om de connectie tussen sterke HI-systemen en sterrenstelsels op  $z = 3$  te bestuderen. We tonen aan dat de sterkst absorberende objecten overeenkomen met lage-massa sterrenstelsels die met de huidige waarneemtechnieken te zwak zijn om gedetecteerd te worden. We laten echter zien dat onze voorspellingen in goede overeenstemming zijn met de bestaande observaties. We wijzen op een sterke anticorrelatie tussen de kolom-

---

dichtheid van sterk absorberende HI-systemen en de loodrechte afstand tussen elk object en het dichtstbijzijnde sterrenstelsel. Bovendien onderzoeken we de correlatie tussen de kolomdichtheid en de verschillende eigenschappen van de sterrenstelsels waar de HI-systemen mee geassocieerd worden.

Net als neutraal waterstof, de voornaamste vorm van brandstof voor stervorming, heeft ook het stof dat aanwezig is in sterrenstelsels een sterke invloed op de stervormingsactiviteit. Het bestuderen van de verdeling en evolutie van stof is daarom ook van groot belang voor ons begrip van de ontwikkeling van sterrenstelsels. Door de lage resolutie van de waarnemingen, die doorgaans op lange golflengte worden gedaan, is de identificatie en het nemen van spectra van individuele infrarode sterrenstelsels in het verre heelal een lastige opgave. Relevante informatie over de evolutie en statistische eigenschappen van deze objecten wordt daarom gehaald uit de oppervlakedichtheid van de bronnen als functie van hun helderheid (m.a.w. het aantal bronnen per oppervlakte met een bepaalde helderheid). In **Hoofdstuk 5** presenteren we een model voor de evolutie van stoffige sterrenstelsels dat gebaseerd is op de waargenomen brondichtheid en roodverschuivingverdeling bij  $850\mu\text{m}$ . We gebruiken een simpel formalisme voor de evolutie van de kleur- en helderheidsverdeling van infrarode stelsels. Met behulp van een nieuw algoritme om de brondichtheden te berekenen onderzoeken we de mate waarin de verschillende vrije parameters van ons model worden beperkt door de beschikbare observationele data. We laten zien dat het model uitstekend in staat is de waargenomen brondichtheid en roodverschuivingverdeling op golflengten van  $70\mu\text{m} \lesssim \lambda \lesssim 1100\mu\text{m}$  te reproduceren en dat het in goede overeenstemming is met de meest recente resultaten van *Herschel* en *SCUBA 2*.



# PUBLICATIONS

1. *Predictions for the relation between strong HI absorbers and galaxies at  $z = 3$*   
Alireza Rahmati & Joop Schaye  
2013, MNRAS, to be submitted.
2. *The effect of recombination radiation on the temperature and ionization state of partially ionized gas*  
Milan Raičević, Andreas H. Pawlik, Joop Schaye & Alireza Rahmati  
2013, MNRAS, submitted.
3. *Stars were born in significantly denser regions in the early Universe*  
Maryam Shirazi, Jarle Brinchmann & Alireza Rahmati  
2013, ApJ, submitted, [arXiv:1307.4758](https://arxiv.org/abs/1307.4758).
4. *The impact of local stellar radiation on the HI column density distribution*  
Alireza Rahmati, Joop Schaye, Andreas H. Pawlik & Milan Raičević  
2013, MNRAS, 431, 2261-2277.
5. *On the evolution of the HI column density distribution in cosmological simulations*  
Alireza Rahmati, Andreas H. Pawlik, Milan Raičević & Joop Schaye  
2013, MNRAS, 430, 2427-2445.
6. *Genesis of the dusty Universe: modeling submillimetre source counts*  
Alireza Rahmati & Paul van der Werf  
2011, MNRAS, 418, 176-194.
7. *New biorthogonal potential–density basis functions*  
Alireza Rahmati & Mir Abbas Jalali  
2009, MNRAS, 393, 1459-1466.



# CURRICULUM VITAE

I was born on 9 February 1982 in Qom, Iran. When I was two years old, I started to taste academic life as my father started his undergraduate studies in Shiraz University. This was the beginning of moving to different parts of the country. I grew up in Qom, Shiraz, Tehran and Yazd, and I went to several different schools in those cities.

In September 2000, I received my pre-university diploma in physics and mathematics from one of the branches of the *National Organization for Development of Exceptional Talents* (NODET) in Yazd, Iran, and started my Physics undergraduate studies at *Amir Kabir University of Technology* (Tehran Polytechnic) in Tehran, Iran. For my B.Sc. thesis, I worked with Prof. dr. B. Maragechi on studying the self-field effects in the dispersion relation of relativistic electron beams. After a very successful performance in national graduate entrance examinations, which brought me the exemption from compulsory military service, I started my Physics graduate studies in September 2004, at *Sharif University of Technology* in Tehran, Iran. For my M.Sc. thesis, I worked with Prof. dr. M. A. Jalali on instabilities of elliptical galaxies, which resulted in my first scientific publication. After finishing my M.Sc. degree in 2006, I was planning to pursue my studies as a Ph.D. student in the US. However, as an Iranian, I was denied a student visa to enter the US and to benefit from my awarded fellowship at Northwestern University. A year later, in September 2007, I started at *Leiden Observatory* as a M.Sc. student, thanks to the Oort and Huygens scholarships. For my minor project, I worked with Prof. dr. M. Franx before starting my M.Sc. thesis with Prof. dr. P. van der Werf on modeling the evolution of dusty galaxies. I graduated cum laude and I decided to stay at Leiden Observatory for my Ph.D., and to work with Prof. dr. Schaye by joining his research group in September 2009.

During my Ph.D., I have presented my work in several institutes and international conferences in Aix-en-Provence (France), Barcelona (Spain), Cambridge, Durham and Edinburgh (UK), Groningen, Dwingeloo (Netherlands), Garching and Heidelberg (Germany), Trieste (Italy), Zurich (Switzerland), Austin, Berkeley, Madison, Santa Barbara, Santa Cruz and San Diego (US).

After a brief working visit at CONICET, Universidad de Buenos Aires (Argentina) in September 2013, I will continue to live my long-standing dream of being an astrophysicist by starting a postdoctoral fellowship at the Max-Planck Institute for Astrophysics (Garching, Germany), in October 2013.





# ACKNOWLEDGEMENTS

With the limited memory that I have, and also because I am not allowed to thank my Ph.D. supervisor here, it is an impossible-to-succeed task to thank properly everyone who helped me to happily finish this thesis and to get to this stage of my life.

I had six years of wonderful time at the Sterrewacht. First, I should thank people who assured me that coming to Leiden was a good idea to begin with: Amir (my M.Sc. supervisor in Iran), Mario Juric and Scott Tremaine (through Mario). Also, I should thank Jarle for helping me to decide to stay. Then, I would like to thank all the past and current sterrewachters whose presence has created one of the best research institutes in the terrestrial world for studying the Universe. In particular, I would like to express my special gratitude to the support staff.

It was a great pleasure to be part of Joop's research group during the past four years. I would like to thank all the past and current group members who helped to establish a solid foundation. Craig, Rob, Marcel, I am glad that I was part of the same group with you. Claudio, we did not have much overlap, but I am grateful for your help whenever I needed it. Andreas, thanks so much for being a great friend, mentor and collaborator. We started our regular and often intense meetings before I officially start my Ph.D. and when you were extremely busy with finishing yours. Thanks for always finding time for me. I am glad that finally we are going to be at the same institute and get to see each other and talk more often. Milan, it was very nice to be your neighbor at the Sterrewacht as I finally had a friend next door to share my cone-related thoughts with. I always remember the good time we had every friday morning in the group meetings enlightening everyone with our cones and getting enlightened in the afternoons! Freeke, I would like to thank you and Olivera for telling me about the life as a member of the group, before I decide to join. Also many thanks for being such a wonderful host during my very short visit at Berkeley. Olivera, it was so nice to have you as a friend and group-mate. Thanks for being worried about me and telling me many secrets about the group and beyond. Joki, I was very happy to see our RT gang growing when you joined the group and I am very sad to leave you alone. I really appreciate your warm friendship and I always have enjoyed your relaxed attitude towards everything. Marcello and Simone, thanks for making the preprint meetings more lively. I always have enjoyed talking to you about science and other things. Alex, Ben, Caroline, Marcel, Marco, Monica and Rob, thanks for being wonderful group-mates during the last couple of years, I will miss you all. And finally, Marijke, I am very happy that you rejoined the group and I am sad that I cannot enjoy your company as a group-mate. I am deeply grateful that you spend your holiday time to help me out with my Nederlandse Samenvatting.

I am grateful for having many friends who have filled my life in the Netherlands with joy. Besides my countless colleagues at the Sterrewacht, I would like to specially thank Azadeh & Saeed, Islam & Nynke, Maryam & Mehdi, Mo-

hammad, Mahsa & Rojman for their priceless friendship. Being with you makes me happy and I am very fortunate to have you as friends.

And finally, my deep gratitude goes to my family for their eternal support. Mom and dad, I am lucky to have such wonderful parents. Thanks for making me who I am. My lovely sisters, Leili and Elahe, thank you for your kindness and your great company whenever we get together. And Maryam, you have always been with me during ups and downs of the last four years. Without you, it would not have been possible to go through this stage of my life so smoothly. I am deeply grateful for your patience as I was often complaining about many things. The most valuable reward I got by being in Leiden during the last four years was your priceless company.

A. Rahmati,  
Buenos Aires, September 2013.

