

Dissertation  
Submitted to the  
Combined Faculties of the Natural Sciences and Mathematics  
of the Ruperto-Carola-University of Heidelberg, Germany  
for the degree of  
Doctor of Natural Sciences

Put forward by

*Daniele Sorini*

*Born in: Trieste, Italy*

*Oral examination: 17 October 2017*



# Constraining the Physics of the Intergalactic and Circumgalactic Media with Lyman- $\alpha$ Absorption

Referees:

Prof. Dr. Joseph F. Hennawi

Prof. Dr. Volker Springel



Topic in German:

Lyman- $\alpha$ -Absorptionsmerkmale, die in Quasarspektren im Rotverschiebungsbereich  $0 < z < 6$  nachgewiesen werden, ermöglichen es, die intergalaktischen und zirkumgalaktischen [circumgalactic] Medien (IGM und CGM) wirksam zu untersuchen, und folglich Modelle der Kosmologie sowie Galaxienentstehung zu bestimmen. Im ersten Teil dieser Doktorarbeit, löse ich einige numerische Herausforderungen, die sich durch kosmologische hydrodynamische Simulationen stellen, indem ich eine neuartige halb-analytische Method entwickle, um verschiedene statistische Eigenschaften der Lyman- $\alpha$  ( $\text{Ly}\alpha$ ) Absorption im IGM mit großen N-Körper Simulationen vorherzusagen. Die in dieser Doktorarbeit entwickelte Method ist genauer als bisherige Versuche in der wissenschaftlichen Literatur. Noch wichtiger ist, dass sie auf Gpc-Skala N-Körper Simulationen angewendet werden kann, was eine detaillierte Erforschung der  $\text{Ly}\alpha$ -Absorption auf beispiellos großen Skalen ermöglicht. Im zweiten Teil der Doktorarbeit betrachte ich Beobachtungen der  $\text{Ly}\alpha$ -Absorption in Spektren von Hintergrundquasaren bei verschiedenen transversalen Abstände von Vordergrundgalaxien (zwischen 25 kpc und 17 Mpc) und vergleiche sie mit den Vorhersagen verschiedener hochmoderner hydrodynamischer kosmologischer Simulationen. Ich zeige zum ersten Mal, dass die Kombination von Beobachtungen der  $\text{Ly}\alpha$ -Absorption im IGM und CGM die in den Simulationen implementierte, aber nicht aufgelöste, Physik (z. B. Feedback) sehr gut eingrenzen kann. Aufgrund der in naher Zukunft geplanten  $\text{Ly}\alpha$ -Absorptionsbeobachtungen, werden die in dieser Doktorarbeit entwickelten Methode die Voraussetzungen für eine noch präzisere Bestimmung der Modelle der Kosmologie und Galaxienentstehung schaffen.

Topic in English:

Lyman- $\alpha$  ( $\text{Ly}\alpha$ ) absorption features detected in quasar spectra in the redshift range  $0 < z < 6$  are a powerful tool to probe the intergalactic and circumgalactic media (IGM and CGM) and, consequently, to constrain models of galaxy formation and cosmology. In the first part of this thesis, I overcome certain numerical challenges posed by cosmological hydrodynamic simulations by developing a novel semi-analytic technique to predict various statistics of the  $\text{Ly}\alpha$  absorption in the IGM with large N-body cosmological simulations. The technique developed in this work is more accurate than previous attempts in the literature. More importantly, it can be applied on Gpc-scale N-body simulations, allowing an accurate investigation of the  $\text{Ly}\alpha$  absorption at unprecedentedly large scales. In the second part of the thesis, I consider observations of  $\text{Ly}\alpha$  absorption in spectra of background quasars at different transverse separations (between 25 kpc and 17 Mpc) from foreground galaxies, and compare them with the predictions of different state-of-the-art hydrodynamic cosmological simulations. For the first time, I show that combining observations of  $\text{Ly}\alpha$  absorption in the IGM and the CGM can tightly constrain the models of sub-resolution physical processes implemented in simulations (e.g., feedback). With near-future high-precision observations of  $\text{Ly}\alpha$  absorption, the tools developed in this work will set the stage for even stronger constraints on models of galaxy formation and cosmology.



RUPERTO-CAROLA-UNIVERSITY OF HEIDELBERG

DOCTORAL THESIS

---

**Constraining the Physics of the  
Intergalactic and Circumgalactic Media  
with Lyman- $\alpha$  Absorption**

---

Author:

Daniele SORINI

Supervisor:

Prof. Dr. Joseph F. HENNAWI

*A thesis submitted in fulfillment of the requirements  
for the degree of Doctor of Natural Sciences*

August 2017



# Declaration of Authorship

I, Daniele SORINI, declare that this thesis titled, 'Constraining the Physics of the Inter-galactic and Circumgalactic Media with Lyman- $\alpha$  Absorption' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

---

Date:

---



RUPERTO-CAROLA-UNIVERSITY OF HEIDELBERG

## *Abstract*

Combined Faculty of Natural Sciences and Mathematics

Department of Physics and Astronomy

Doctor of Natural Sciences

### **Constraining the Physics of the Intergalactic and Circumgalactic Media with Lyman- $\alpha$ Absorption**

by Daniele SORINI

Lyman- $\alpha$  ( $\text{Ly}\alpha$ ) absorption features detected in quasar spectra in the redshift range  $0 < z < 6$  are a powerful tool to probe the intergalactic and circumgalactic media (IGM and CGM) and, consequently, to constrain models of galaxy formation and cosmology. In the first part of this thesis, I overcome certain numerical challenges posed by cosmological hydrodynamic simulations by developing a novel semi-analytic technique to predict various statistics of the  $\text{Ly}\alpha$  absorption in the IGM with large N-body cosmological simulations. The technique developed in this work is more accurate than previous attempts in the literature. More importantly, it can be applied on Gpc-scale N-body simulations, allowing an accurate investigation of the  $\text{Ly}\alpha$  absorption at unprecedentedly large scales. In the second part of the thesis, I consider observations of  $\text{Ly}\alpha$  absorption in spectra of background quasars at different transverse separations (between 25 kpc and 17 Mpc) from foreground galaxies, and compare them with the predictions of different state-of-the-art hydrodynamic cosmological simulations. For the first time, I show that combining observations of  $\text{Ly}\alpha$  absorption in the IGM and the CGM can tightly constrain the models of sub-resolution physical processes implemented in simulations (e.g., feedback). With near-future high-precision observations of  $\text{Ly}\alpha$  absorption, the tools developed in this work will set the stage for even stronger constraints on models of galaxy formation and cosmology.



RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG

## *Zusammenfassung*

Naturwissenschaftlich-Matematischen Gesamtfakultät  
Fakultät für Physik und Astronomie

Doctor Rerum Naturalium

### **Bestimmung der Physik der intergalaktischen und zirkumgalaktischen Medien mit Hilfe von Lyman- $\alpha$ Absorption**

von Daniele SORINI

Lyman- $\alpha$ -Absorptionsmerkmale, die in Quasarspektren im Rotverschiebungsbereich  $0 < z < 6$  nachgewiesen werden, ermöglichen es, die intergalaktischen und zirkumgalaktischen [circumgalactic] Medien (IGM und CGM) wirksam zu untersuchen, und folglich Modelle der Kosmologie sowie Galaxienentstehung zu bestimmen. Im ersten Teil dieser Doktorarbeit, löse ich einige numerische Herausforderungen, die sich durch kosmologische hydrodynamische Simulationen stellen, indem ich eine neuartige halb-analytische Method entwickle, um verschiedene statistische Eigenschaften der Lyman- $\alpha$  ( $\text{Ly}\alpha$ ) Absorption im IGM mit großen N-Körper Simulationen vorherzusagen. Die in dieser Doktorarbeit entwickelte Method ist genauer als bisherige Versuche in der wissenschaftlichen Literatur. Noch wichtiger ist, dass sie auf Gpc-Skala N-Körper Simulationen angewendet werden kann, was eine detaillierte Erforschung der  $\text{Ly}\alpha$  Absorption auf beispiellos großen Skalen ermöglicht. Im zweiten Teil der Doktorarbeit betrachte ich Beobachtungen der  $\text{Ly}\alpha$ -Absorption in Spektren von Hintergrundquasaren bei verschiedenen transversalen Abstände von Vordergrundgalaxien (zwischen 25 kpc und 17 Mpc) und vergleiche sie mit den Vorhersagen verschiedener hochmoderner hydrodynamischer kosmologischer Simulationen. Ich zeige zum ersten Mal, dass die Kombination von Beobachtungen der  $\text{Ly}\alpha$ -Absorption im IGM und CGM die in den Simulationen implementierte, aber nicht aufgelöste, Physik (z. B. Feedback) sehr gut eingrenzen kann. Aufgrund der in naher Zukunft geplanten  $\text{Ly}\alpha$ -Absorptionsbeobachtungen, werden die in dieser Doktorarbeit entwickelten Methode die Voraussetzungen für eine noch präzisere Bestimmung der Modelle der Kosmologie und Galaxienentstehung schaffen.



*In memoria dei miei nonni*  
*In memory of my grandfathers*



# Contents

<b>Declaration of Authorship</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>Zusammenfassung</b>	<b>iii</b>
<b>Contents</b>	<b>iv</b>
<b>List of Figures</b>	<b>viii</b>
<b>List of Tables</b>	<b>x</b>
<b>Abbreviations</b>	<b>xi</b>
<b>Symbols and Conventions</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Observational Features of the Ly $\alpha$ Absorption Line . . . . .	6
1.2 Ly $\alpha$ Absorption in the IGM and the CGM . . . . .	7
1.3 Thesis Outline . . . . .	9
<b>2 Modeling the Ly<math>\alpha</math> Forest in Collisionless Simulations</b>	<b>13</b>
2.1 Simulations . . . . .	16
2.1.1 Ly $\alpha$ Skewers . . . . .	19
2.2 Limitations of Approximate Methods . . . . .	19
2.2.1 Gaussian Smoothing . . . . .	20
2.2.2 Fluctuating Gunn-Peterson Approximation . . . . .	20
2.2.3 Accuracy of FGPA . . . . .	22
2.3 Iteratively Matched Statistics . . . . .	27
2.3.1 3D Iteratively Matched Statistics . . . . .	27
2.3.2 1D Iteratively Matched Statistics . . . . .	31
2.4 Validation of Iteratively Matched Statistics . . . . .	31
2.4.1 Skewers . . . . .	33
2.4.2 Comparison of the Methods . . . . .	34
2.4.3 Accuracy versus Smoothing Length . . . . .	38

2.5	Large-Volume Collisionless Simulation . . . . .	42
2.6	Comparison to Previous Works . . . . .	46
2.7	Chapter Summary . . . . .	48
<b>3</b>	<b>Constraining Simulations with Ly<math>\alpha</math> Absorption Profiles around Galaxies</b>	<b>51</b>
3.1	Simulations . . . . .	54
3.1.1	Nyx . . . . .	54
3.1.2	Illustris . . . . .	55
3.2	Modeling . . . . .	56
3.2.1	Selection of Halos . . . . .	56
3.2.2	Construction of Samples of Skewers . . . . .	58
3.2.3	Simulating HI Absorption . . . . .	58
3.2.4	Altering the Temperature of the CGM . . . . .	59
3.3	Results . . . . .	63
3.3.1	Simulated Spectra . . . . .	63
3.3.2	Comparison with Observations . . . . .	65
3.3.2.1	Quasar Hosts . . . . .	65
3.3.2.2	Damped Ly $\alpha$ Absorbers . . . . .	68
3.3.2.3	Mean Ly $\alpha$ Transmission Profile around LBGs . . . . .	70
3.3.2.4	Median Ly $\alpha$ Optical Depth around LBGs . . . . .	72
3.3.3	Summary . . . . .	73
3.4	Discussion . . . . .	74
3.4.1	Temperature and Density of the Gas . . . . .	74
3.4.1.1	Temperature-Density Relationship in the IGM . . . . .	75
3.4.1.2	Radial Temperature and Density Profiles . . . . .	76
3.4.1.3	Temperature-Density Relationship in the CGM . . . . .	80
3.4.1.4	Summary . . . . .	82
3.4.2	Halo Mass . . . . .	85
3.4.3	Redshift of Foreground Objects . . . . .	86
3.4.4	Sample Size of Observed Spectra . . . . .	87
3.4.5	Sub-resolution Physics . . . . .	89
3.4.6	Comparison with Previous Work . . . . .	90
3.5	Conclusions and Perspectives . . . . .	91
<b>4</b>	<b>Testing Cosmological Predictions of Kinetic Field Theory</b>	<b>95</b>
4.1	An Overview of Kinetic Field Theory . . . . .	99
4.2	Millennium Run . . . . .	100
4.3	Statistics Considered . . . . .	101
4.4	Computing Power Spectra From Simulations . . . . .	102
4.5	Discussion and Outlook . . . . .	107
<b>5</b>	<b>Conclusions and Perspectives</b>	<b>109</b>
<b>A</b>	<b>Optimal Smoothing Length for Velocity</b>	<b>117</b>

---

<b>B Calibration of Halo Masses</b>	<b>120</b>
<b>C Generating Mock Spectra from a Moving-Mesh Code</b>	<b>123</b>
<b>D From Cross-Correlation to Mean Flux Profile</b>	<b>124</b>
<b>E Mean Flux Fluctuations at Small Separations from DLAs</b>	<b>128</b>
<b>F Outflows in the CGM</b>	<b>129</b>
<b>G Convergence of Power Spectra</b>	<b>133</b>
<b>List of Author's Publications</b>	<b>136</b>
<b>Bibliography</b>	<b>137</b>
<b>Acknowledgements</b>	<b>159</b>

# List of Figures

1.1	Cosmic Web . . . . .	3
1.2	Ly $\alpha$ Features in QSO Absorption Spectrum . . . . .	8
2.1	Resolution of Nyx . . . . .	18
2.2	Accuracy of FGPA: Skewers . . . . .	23
2.3	Accuracy of FGPA: Line-of-Sight Power Spectrum and PDF . . . . .	24
2.4	Accuracy of FGPA: 3D Power Spectrum . . . . .	25
2.5	Conceptual Map of the Methods Tested . . . . .	29
2.6	Accuracy of IMS: Skewers . . . . .	32
2.7	Accuracy of IMS: Line-of-Sight Power Spectrum and PDF . . . . .	34
2.8	Accuracy of IMS: 3D Power Spectrum . . . . .	35
2.9	Accuracy versus Smoothing Length: Line-of-Sight Power Spectrum and PDF . . . . .	36
2.10	Accuracy versus Smoothing Length: 3D Power Spectrum . . . . .	37
2.11	Application to Large N-Body Simulation: Skewers . . . . .	41
2.12	Application to Large N-Body Simulation: Line-of-Sight Power Spectrum and PDF . . . . .	42
2.13	Application to Large N-Body Simulation: 3D Power Spectrum . . . . .	43
3.1	Temperature and Density Slices around Halos . . . . .	60
3.2	Test of Semi-Analytic Method . . . . .	61
3.3	Examples of Simulated Skewers . . . . .	64
3.4	Mean Ly $\alpha$ Flux Fluctuations around QSOs . . . . .	66
3.5	Mean Ly $\alpha$ Flux Fluctuations around DLAs . . . . .	69
3.6	Mean Ly $\alpha$ Flux Fluctuations and Median Ly $\alpha$ Optical Depth around LBGs . . . . .	71
3.7	IGM Temperature-Density Relationship in Simulations . . . . .	75
3.8	Radial Temperature and Density Profiles around QSO Hosts . . . . .	77
3.9	Radial Temperature and Density Profiles around LBG/DLA Hosts . . . . .	78
3.10	Temperature-Density Relationship at Different Distances from QSO Hosts . . . . .	83
3.11	Temperature-Density Relationship at Different Distances from LBG/DLA Hosts . . . . .	84
3.12	Scatter in the Mean Ly $\alpha$ Flux around QSOs and DLAs due to Size of Observed Sample . . . . .	87
3.13	Scatter in the Median Ly $\alpha$ Optical Depth due to Size of Observed Sample . . . . .	88
4.1	Testing Computation of Power Spectrum of Density Fluctuations . . . . .	103
4.2	Accuracy of KFT Prediction of Power Spectrum of Density Fluctuations . . . . .	104
4.3	Power Spectra of Momentum Density Fluctuations Predicted by Millennium Run . . . . .	106

---

A.1	Effect of Smoothing the Velocity Field: Line-of-Sight Power Spectrum and PDF . . . . .	118
A.2	Effect of Smoothing the Velocity Field: 3D Power Spectrum . . . . .	119
B.1	Calibration of Halo Masses in Simulation . . . . .	122
F.1	Radial Velocity-Density Relationship at Different Distances from QSO Hosts . . . . .	130
F.2	Radial Velocity-Density Relationship at Different Distances from DLA Hosts . . . . .	131
G.1	Convergence of Power Spectrum of Density Fluctuations . . . . .	133
G.2	Convergence of Power Spectra of Momentum Density Fluctuations . . . . .	135

# List of Tables

2.1	Inputs of Methods . . . . .	27
3.1	Parameters Adopted to Reproduce Observations from Simulations . . . . .	57
D.1	Mean Ly $\alpha$ Flux at Large Impact Parameter from DLAs . . . . .	127
D.2	Mean Ly $\alpha$ Flux at Large Impact Parameter from QSOs . . . . .	127
E.1	Mean Ly $\alpha$ Flux at Small Impact Parameter from DLAs . . . . .	128

# Abbreviations

All abbreviations and acronyms are defined at their first appearance in the main text. If they are used more than once, they are also listed in the following table, in alphabetic order.

<b>AGN</b>	<b>A</b> ctive <b>G</b> alactic <b>N</b> ucleus/ <i>i</i>
<b>BAO</b>	<b>B</b> aryon <b>A</b> coustic <b>O</b> scillations
<b>BOSS</b>	<b>B</b> aryon <b>O</b> scillation <b>S</b> pectroscopic <b>S</b> urvey
<b>CDDF</b>	<b>C</b> olumn <b>D</b> ensity <b>D</b> istribution <b>F</b> unction
<b>CIC</b>	<b>C</b> loud- <b>I</b> n- <b>C</b> ell
<b>CGM</b>	<b>C</b> ircum <b>G</b> alactic <b>M</b> edium
<b>CLAMATO</b>	<b>C</b> OSMOS <b>L</b> yman <b>A</b> lpha <b>M</b> apping <b>A</b> nd <b>T</b> omography <b>O</b> bservations
<b>CMB</b>	<b>C</b> osmic <b>M</b> icrowave <b>B</b> ackground
<b>DESI</b>	<b>D</b> ark <b>E</b> nergy <b>S</b> pectroscopic <b>I</b> nstrument
<b>DLA</b>	<b>D</b> amped <b>L</b> yman- $\alpha$ <b>A</b> bsorber
<b>DM</b>	<b>D</b> ark <b>M</b> atter
<b>FGPA</b>	<b>F</b> luctuating <b>G</b> unn- <b>P</b> eterson <b>A</b> pproximation
<b>GS+FGPA</b>	<b>G</b> aussian <b>S</b> moothering + <b>F</b> luctuating <b>G</b> unn- <b>P</b> eterson <b>A</b> pproximation
<b>HPM</b>	<b>H</b> ydro <b>P</b> article <b>M</b> esh
<b>IGM</b>	<b>I</b> nter <b>G</b> alactic <b>M</b> edium
<b>IMS</b>	<b>I</b> teratively <b>M</b> atched <b>S</b> tatistics
<b>KBSS</b>	<b>K</b> eck <b>B</b> aryonic <b>S</b> tructure <b>S</b> urvey
<b>KFT</b>	<b>K</b> inetic <b>F</b> ield <b>T</b> heory
$\Lambda$ CDM	$\Lambda$ - <b>C</b> old <b>D</b> ark <b>M</b> atter
<b>LBG</b>	<b>L</b> yman <b>B</b> reak <b>G</b> alaxy
$\text{Ly}\alpha$	<b>L</b> yman- $\alpha$
<b>PDF</b>	<b>P</b> robability <b>D</b> ensity <b>F</b> unction

---

<b>PM</b>	<b>P</b> article <b>M</b> esh
<b>QSO</b>	<b>Q</b> uasi <b>S</b> tellar <b>O</b> bject (quasar)
<b>SPH</b>	<b>S</b> moothed <b>P</b> article <b>H</b> ydrodynamics
<b>UV</b>	<b>U</b> ltra <b>V</b> iolet
<b>UVB</b>	<b>U</b> ltra <b>V</b> iolet <b>B</b> ackground
<b>WMAP</b>	<b>W</b> ilkinson <b>M</b> icrowave <b>A</b> nisotropy <b>P</b> robe
<b>1DPS</b>	<b>1</b> - <b>D</b> imensional (line-of-sight) <b>P</b> ower <b>S</b> pectrum
<b>3DPS</b>	<b>3</b> - <b>D</b> imensional <b>P</b> ower <b>S</b> pectrum
<b>4MOST</b>	<b>4</b> -metre <b>M</b> ulti- <b>O</b> bject <b>S</b> pectroscopic <b>T</b> elescope

# Symbols and Conventions

- Ionization states** If a certain element is ionized  $n$  times, its chemical symbol is followed by  $n+1$ , written as a Roman number. For example, HI denotes neutral hydrogen, while HII denotes ionized hydrogen.
- Units of distance** Throughout the thesis, co-moving units are denoted with a “c” in front of the unit of measure (i.e. ckpc, cMpc, cGpc). Otherwise, distances are expressed in physical units (i.e. kpc, Mpc, Gpc).
- Vectors** A three-dimensional vector is indicated boldface type, e.g.  $\mathbf{x}$ . In this case,  $x$  indicates its norm, i.e.  $x \equiv |\mathbf{x}| = \sqrt{|\mathbf{x} \cdot \mathbf{x}|}$ .
- Fourier transform** Let  $x \in \mathbb{R}^n$  and  $f(x)$  be a scalar function. The Fourier transform  $\hat{f}(k)$  is defined by
$$\hat{f}(k) = \int_{\mathbb{R}^n} d^n x e^{-ik \cdot x} f(x).$$
To make notation lighter, I shall conform to the widely spread convention of omitting the hat when denoting the Fourier transform.
- Dirac delta** Let  $x \in \mathbb{R}^n$ . I denote the Dirac delta as  $\delta_D(x)$ , regardless the dimension  $n$ . It will be easily inferred from context.



*“Io stimo più il trovar un vero, benché di cosa leggiera, che ’l disputar lungamente delle massime questioni, senza conseguir verità nissuna.”*

Galileo Galilei

*“I value more finding some truth, albeit in a little subject, than lengthily debating the highest matters, without achieving any truth.”*

Galileo Galilei



# Chapter 1

## Introduction

With the discovery of the first extragalactic nebulae (Hubble, 1925), it became clear that the Universe is not limited to our Galaxy. Since then, more than nine decades of observational evidence and theoretical efforts led to the establishment of the  $\Lambda$ -Cold Dark Matter ( $\Lambda$ CDM) model as the standard paradigm for structure formation and evolution in a cosmological context (Dodelson, 2003, Fukugita et al., 1998, Padmanabhan, 2006, and references therein). Within this framework, the content of the Universe can be divided into three main components: baryons, dark matter (DM) and dark energy. The most recent estimates of their relative contributions to the total energy density of the Universe are 0.05, 0.26 and 0.69, respectively (Planck Collaboration et al., 2016). In the astrophysical lexicon, the term *baryons* refers to all matter made of particles described by the Standard Model of particle physics. Dark matter is a type of matter that neither emits nor absorbs light, and interacts mainly or solely through gravitation. While indirect evidence for DM is provided by the mismatch between luminosity and mass distribution in the Milky Way (Oort, 1932), other galaxies (Rubin et al., 1980), and galaxy clusters (Zwicky, 1933), there is no direct detection of DM at present. The nature of DM is highly debated, and represents one of the fundamental questions in both astrophysics and particle physics<sup>1</sup>. The nature of dark energy is even less clear, despite being the dominant energy component of the Universe. In the  $\Lambda$ CDM model, it

---

<sup>1</sup>Within the  $\Lambda$ CDM model, DM is described as cold dark matter (CDM), i.e. non-relativistic self-gravitating particles (Blumenthal et al., 1984, Occhionero et al., 1984, Primack, 1984). Possible candidates for CDM are axions, massive compact halo objects (MACHOs) and weak interacting massive particles (WIMPs) (Arun et al., 2017, Peter, 2012, for recent reviews). Unless otherwise indicated, in this thesis I shall consider CDM when referring to DM. Some alternative models of DM are hot dark matter (HDM), describing DM as ultra-relativistic particles (e.g., neutrinos), or warm dark matter (WDM), within which DM presents intermediate characteristics between CDM and HDM (Arun et al., 2017, for a recent review). Sterile neutrinos are possible candidates for WDM (Dodelson & Widrow, 1993, Widrow, 1993). Other examples of alternative models of DM are self-interacting dark matter (SIDM), meta-cold dark matter (mCDM) and fuzzy cold dark matter (Arun et al., 2017, for a recent review).

is described as a cosmological constant, and is responsible for the accelerated expansion of the Universe (Perlmutter et al., 1999, Riess & *et al.*, 1998, Schmidt et al., 1998).<sup>2</sup>

According to the  $\Lambda$ CDM paradigm, the Universe began with the Big Bang singularity, which was followed by an era of accelerated expansion, called Inflation (Barrow & Turner, 1981, Guth, 1981). During Inflation, tiny density perturbations were seeded in the matter density of the Universe, which was otherwise homogeneous. Approximately three minutes after the Big Bang, the baryonic component of the Universe was in the state of a hot ionized plasma, composed mainly of electrons, neutrinos and light nuclei ( $\sim 75\%$  hydrogen,  $\sim 25\%$  helium, and traces of their isotopes and other heavier elements), that was tightly coupled to photons. As the Universe kept expanding, it cooled down, allowing photons to decouple from baryons and form the cosmic microwave background (CMB). In the same epoch (redshift<sup>3</sup>  $z \approx 1100$ ), electrons and primordial nuclei recombined, forming the first atoms. The detection of CMB photons allowed mapping the temperature fluctuations of the Universe at the epoch of recombination (Planck Collaboration et al., 2016); since one cannot detect photons carrying information from earlier epochs, the CMB is essentially the earliest currently available image of the Universe. After recombination, the Universe became neutral and transparent to radiation, entering the so-called “Cosmic Dark Ages”.

Throughout the Cosmic Dark Ages, the matter density perturbations kept growing due to gravitational instability, giving rise to the first structures. Dark matter was the main actor in setting the gravitational potential wells, within which it could cluster together with baryons. Hence, collapsing matter started forming halos in the densest regions, connected through a network of filaments. This characteristic configuration of the matter distribution in the Universe is called “cosmic web” (see the upper left and right panels of Figure 1.1).

Within halos, the emergence of the accretion shock split the gas into two components: a low-density supersonic flow, and a high-density subsonic flow (White & Rees, 1978). The latter got shock heated to the virial temperature<sup>4</sup>, beginning the accretion of gas

<sup>2</sup>The cosmological constant ( $\Lambda$ ) is a form of energy with negative pressure and constant density throughout the history of the Universe (Einstein, 1917). Some models outside the  $\Lambda$ CDM paradigm consider a time-varying dark energy (e.g., Khurshudyan & Khurshudyan, 2017, and references therein). Other theories describe dark energy as a scalar field, representing the fifth fundamental force of nature (“quintessence”, see, e.g., Caldwell et al., 1998, Steinhardt & Caldwell, 1998), or aim at treating DM and dark energy as two aspects of the same phenomenon (interacting dark energy; see, e.g., Khurshudyan & Khurshudyan, 2017, and references therein). An alternative approach consists in eliminating the need for dark energy by describing the gravitational interaction with modified gravity theories (see, e.g., Amendola & Tsujikawa, 2010, for a review) rather than general relativity (Einstein, 1915a,b,c, 1916).

<sup>3</sup>Due to the expansion of the Universe, a photon emitted with wavelength  $\lambda_e$  at the time  $t_e$  is detected at a larger wavelength  $\lambda_o$  by an observer at time  $t_o$ . This effect is measured by the redshift  $z$ , defined as  $1 + z = \lambda_o/\lambda_e = a_o/a_e$ , where  $a_o$  and  $a_e$  are the scale factors at times  $t_o$  and  $t_e$ , respectively.

<sup>4</sup>The virial theorem states that, for a stable, self-gravitating ensemble of gas particles, the total kinetic energy  $K$  and the potential energy of the system  $U$  are related by the equality  $2K + U = 0$ . The exact expression of the virial temperature of the ensemble depends on the degrees of freedom of the gas particles (see, e.g., Binney & Tremaine, 2008).

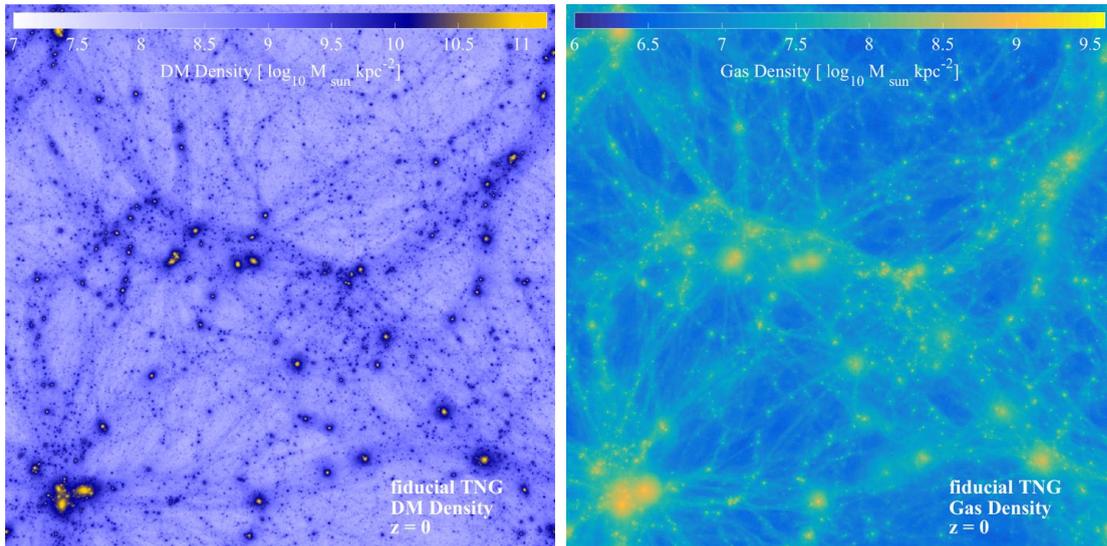
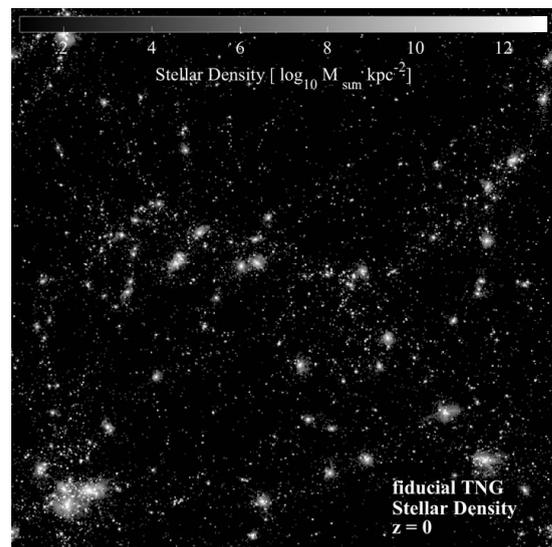


FIGURE 1.1: Projections of DM (upper left panel), baryon (upper right panel), and stellar mass density (lower right panel), on a side of the  $(37 \text{ cMpc})^3$  box simulated with the fiducial IllustrisTNG run, at  $z = 0$ . The cosmic web can be clearly seen in the left and middle panels. At low densities, baryon density represents a smoother version of the corresponding DM density, because, unlike DM, the gravitational collapse of baryons is counteracted by the pressure of gas (see text for details). The right panel shows that star formation occurs within DM halos. Figure adapted from Pillepich et al. (2017).



within halos. Subsequently, the accreting gas cooled down due to bremsstrahlung, radiative recombination, collisional ionization and collisional excitation (Meiksin, 2009). At the same time, photons kept heating the gas through photo-ionization and photo-heating (Cantalupo, 2010, Gnedin & Hollon, 2012, Kannan et al., 2014). Depending on the magnitude of the net cooling rate, set by the balance between the aforementioned cooling and heating processes, the cooling time of the gas can be larger or smaller than the free-fall time (Rees & Ostriker, 1977, White & Frenk, 1991, White & Rees, 1978). In the former case, the gas would reach a pressure-supported hydrostatic equilibrium configuration, contracting slowly. In the latter case, the gas would cool fast, being able to form stars efficiently (see the lower right panel of Figure 1.1).

It is not yet precisely clear when star formation commenced. It was a continuous process that likely started at  $z \approx 50$  (Abel et al. 2002, Gao et al. 2007, Yoshida et al. 2006; but see also Naoz et al. 2006) and led to the formation of the first galaxies by  $z \approx 10$  (Benson,

2010, for a review). The star formation rate peaked in the redshift range  $1 \lesssim z \lesssim 3$ , and is still ongoing (Madau & Dickinson, 2014, for a review). Apart from stars, gas accretion could also form super-massive black holes in the central region of massive galaxies. These black holes could power active galactic nuclei (AGN), i.e. compact regions characterized by a strong emission of non-stellar radiation that include quasars (QSOs). The AGN represent an extra source of heating, which competes with star formation. In fact, various models propose that star formation and/or black-hole accretion could be self-regulating mechanisms (Fabian, 1999, King, 2003, Silk & Rees, 1998).

Parallel to star formation within halos, intergalactic gas (called intergalactic medium, IGM) underwent some important phase transitions. Via the emission of ultraviolet (UV) photons, newly formed structures began to reionize the IGM, and the Universe entered the Epoch of Reionization. According to the standard picture, stellar radiation from galaxies was the main driver of the reionization of the hydrogen atoms present in the IGM (Becker & Bolton, 2013, Faucher-Giguère et al., 2008b, Madau et al., 1999, Shapiro et al., 1994), although some recent findings suggest that QSOs may have contributed to a larger extent than previously thought (Chardin et al., 2015, Giallongo et al., 2015, Madau & Haardt, 2015). The background of UV photons (UVB) in the IGM reionized neutral helium ( $\text{HeI} \rightarrow \text{HeII}$ ) and singly-ionized helium ( $\text{HeII} \rightarrow \text{HeIII}$ ), too. The general wisdom is that QSOs dominated the latter transition, since they supplied enough energetic photons by  $z \approx 3$  (Furlanetto & Oh 2008, Wyithe & Loeb 2003; but see also Worseck et al. 2016). On the contrary, it is unlikely that stars could produce enough photons for the second reionization of helium (Bromm & Yoshida, 2011, Venkatesan et al., 2003). Other models of reionization invoke more exotic sources of photons (X-rays from X-rays binaries Furlanetto et al. 2006, Mirabel et al. 2011, supernovae shocks Johnson & Khochfar 2011, supernovae-accelerated electronic cosmic rays Oh 2001, DM annihilations Belikov & Hooper 2009; see also Dijkstra et al. 2004, McQuinn 2012). According to the current constraints, the reionization of hydrogen was completed by  $z \approx 6$  (Fan et al., 2006, McGreer et al., 2015), while helium was fully reionized in the redshift range  $2 < z < 2.7$  (e.g., Shull et al. 2010; see McQuinn 2016, for a review).

From the previous overview of the history of the Universe, it emerges that the  $\Lambda$ CDM model successfully explains how it transitioned from an initial quasi homogeneous state to the large-scale structures that can be observed today. The various phase transitions of the Universe and star formation history can be described within the  $\Lambda$ CDM framework as well (e.g., Springel & Hernquist, 2003). However, there are still several aspects to clarify, especially as far as the Epoch of Reionization and galaxy formation are concerned. For example, as briefly mentioned earlier, the beginning of reionization (which is tied to primordial star formation), as well as the nature of the sources that powered it, is still under debate (McQuinn, 2016). Moreover, the morphology of reionization depends on the primary sources of photons: if dominated by rare and luminous objects like QSOs, reionization would be more patchy than galaxy-driven reionization (Meiksin, 2009). Regarding galaxy formation, while state-of-the-art cosmological hydrodynamic simulations

(e.g., Pillepich et al., 2017, Schaye et al., 2015, Vogelsberger et al., 2014b) successfully reproduce a plethora of observables (e.g., the star formation efficiency Behroozi et al. 2013, Guo et al. 2011, Moster et al. 2013, the evolution of the star formation rate density Behroozi et al. 2013, Oesch et al. 2015, the black-hole-stellar-mass relationship within galaxies Kormendy & Ho 2013, McConnell & Ma 2013, the gas fraction within halos Giodini et al. 2009, Lovisari et al. 2015, the stellar mass function Baldry et al. 2012, 2008, Bernardi et al. 2013, D’Souza et al. 2015, and the stellar half-mass radii of galaxies Baldry et al. 2012, Shen et al. 2003), it is still challenging to precisely model the physics governing star formation, galactic winds, gas accretion, supernovae explosion and super-massive black hole accretion and merging. Indeed, these processes are often not resolved in simulations, consequently they need to be implemented through certain feedback prescriptions (e.g., Springel 2000, Springel et al. 2005a; see Heckman & Thompson 2017, for a review).

A greater insight both in the epoch of reionization and galaxy formation can be achieved by exploiting some observational features of the residual neutral hydrogen in the cosmic web. While mostly made of ionized gas, the IGM still retains a small fraction of HI ( $\sim 10^{-5}$  at  $z \approx 2$ , see, e.g., Meiksin 2009) that imprints a characteristic pattern, known as “Lyman- $\alpha$  Forest”, in the absorption spectra of QSOs. Such absorption features can be used to map the distribution of HI in the Universe at different redshifts, thus providing us with a better understanding of the reionization history. The IGM also represents an extraordinary cosmological probe, which is able to trace density fluctuations in the redshift range  $0 \lesssim z \lesssim 6$  (see Meiksin, 2009, for a review). As such, it allows testing cosmological models and reconstructing the large-scale distribution of matter. In addition, the IGM contains 80% of baryons at redshift  $z \gtrsim 1.5$  (Prochaska & Hennawi 2009; see also Meiksin 2009, Rauch 1998, and references therein), acting as a reservoir for forming galaxies. Consequently, the IGM is connected with galaxy formation and evolution, too.

At the interface between the IGM and galaxies one can define another environment, the circumgalactic medium (CGM). The CGM is generally considered to be the region within 300 kpc from the center a galaxy. If a galaxy lies at a small transverse separation from a background QSO, one can observe Lyman- $\alpha$  ( $\text{Ly}\alpha$ ) absorption due to the neutral hydrogen present in the CGM.  $\text{Ly}\alpha$  absorption is then a useful tool to probe galactic physics, too. Indeed, the absorption features can provide us with information about the temperature, composition and clumpiness of the CGM (Barnes et al., 2014, for a review). This helps shedding light on the physical processes fundamental to galaxy formation that occur within such medium, e.g., gas accretion and outflows.

Clearly, the  $\text{Ly}\alpha$  absorption line represents a powerful tool for astrophysics and cosmology, as it allows probing a wide range of scales, from hundreds of Mpc down to galactic scales. In § 1.1 I discuss the main observational features of the this line, as well as the underlying physics. In § 1.2 I introduce the basic physical processes that need to be included in simulations to model  $\text{Ly}\alpha$  absorption in the IGM and the CGM, leaving a

more detailed introduction to this subject to the relevant chapters. Finally, in § 1.3, I illustrate the outline of this thesis.

## 1.1 Observational Features of the Ly $\alpha$ Absorption Line

Nearly coincident with the discovery of the CMB and shortly after the discovery of the first QSO (Hazard et al., 1963, Schmidt, 1963), Gunn & Peterson (1965) detected certain absorption features in its spectrum, proving that the cosmic mass density of HI was much smaller than the spatially averaged hydrogen of all stars in the Universe. Assuming that the Big Bang theory was correct, this meant that either galaxy formation was an extremely efficient process, exploiting almost all the available neutral hydrogen in intergalactic space, or that a great fraction of the gas was ionized.

Since then, almost all of our knowledge about the IGM derives from observations in the optical and UV bands (Meiksin, 2009). These observations have mainly concerned absorption spectra of QSOs, but also IGM absorption features in the spectra of gamma ray bursts (GRBs; Totani et al. 2006). Shortly after the discovery by Gunn & Peterson (1965), it was clear that the absorption features could be exploited as a cosmological probe. In fact, it was realized that such features should have appeared from neutral hydrogen in cosmological structures (Bahcall & Salpeter, 1965, Wagoner, 1967). The presence of the absorption lines turned out to be very common in several spectra, and Lynds (1971) realized that they corresponded to Ly $\alpha$  transitions. Sargent et al. (1980) showed that the widths of the lines suggested gas temperatures of around  $10^4$  K, corresponding to the order of magnitude required for photoionizing a gas of primordial composition, i.e. with low metal<sup>5</sup> contamination.

To understand the physics underlying the Ly $\alpha$  absorption features in QSO spectra, one first of all needs to consider the radiation emitted by these objects, which spans different regions of the electromagnetic spectrum, from gamma rays to the far infrared (see, e.g., Kembhavi & Narlikar, 1999). About 10% of QSOs emit radio frequencies and many of them show extensive UV emission as well (see, e.g., Kembhavi & Narlikar, 1999), which can include a prominent Ly $\alpha$  emission peak. The UV radiation can be absorbed by the neutral hydrogen present in the IGM spread between the observer and the QSO. Indeed, the Ly $\alpha$  line of HI, corresponding to the transition of the electron between the first and second energy levels, is a resonant line with rest-frame wavelength  $\lambda_{\text{Ly}\alpha} = 1216 \text{ \AA}$ . Therefore, whenever a photon of this wavelength encounters an atom of HI in the ground state, there is a high probability that it is absorbed, promoting the electron to the second energy level. The electron will eventually go back to the fundamental level, re-emitting a Ly $\alpha$  photon. However, in general the outgoing photon will be emitted in a direction different from the one of the incoming photon. That is why the neutral hydrogen in

<sup>5</sup>In the astrophysical lexicon, *metals* comprehends all elements other than hydrogen and helium (and their isotopes).

the intervening IGM effectively prevents part of the Ly $\alpha$  photons emitted by the QSO from reaching the observer. As a result, an absorption line in the QSO spectrum will be observed.

As a photon travels from the QSO to the observer through the IGM, it is continuously redshifted due to the expansion of the Universe. Hence, atoms of neutral hydrogen at different locations along the line of sight will “see” the incoming photon at different wavelengths. If  $z_Q$  is the redshift of the QSO and  $z$  an intermediate redshift between the observer and the QSO itself, a photon with rest-frame wavelength  $\lambda$  at  $z_Q$  will have a wavelength equal to  $\lambda a(z)/a(z_Q)$  at redshift  $z$ . Therefore, the photons emitted in the range of wavelengths  $a(z_Q)/a(z)\lambda_{\text{Ly}\alpha} < \lambda < \lambda_{\text{Ly}\alpha}$  can be absorbed by the IGM between redshift  $z$  and  $z_Q$ . This effect gives rise to an attenuation of the electromagnetic radiation detected in the spectrum. Furthermore, when an atom at redshift  $z$  undergoes a Lyman- $\alpha$  transition, one will not observe a dip in the spectrum at  $\lambda_{\text{Ly}\alpha}$ , but at  $(1+z)\lambda_{\text{Ly}\alpha}$ . Hence, when observing a QSO spectrum, one will not see one unique absorption line, but many replicas of the same line in the range of wavelengths  $\lambda_{\text{Ly}\alpha} < \lambda < (1+z_Q)\lambda_{\text{Ly}\alpha}$ , corresponding to the absorption by neutral hydrogen atoms at different redshifts. This collection of spectral features is called “Ly $\alpha$  forest”. As an example, Figure 1.2 shows the spectrum of the quasar HS0741+4741 at  $z = 3.22$  (Songaila, 2006). The peak at 5130Å is the Ly $\alpha$  emission of the QSO. Blueward of the peak, one can clearly see the Ly $\alpha$  forest.

The Ly $\alpha$  forest has been studied with growing interest since its discovery. Given that the Ly $\alpha$  forest traces the underlying distribution of HI, which in turn relates to the density fluctuations in the distribution of baryons, the statistical properties of the Ly $\alpha$  forest could be modeled directly from first principles. With the development of cosmological simulations, it became possible to compute the density fluctuations at a certain redshift starting from initial conditions suggested by cosmological models and predict the statistics of the absorption lines. In the past two decades, an ever-growing effort has been devoted to constraining cosmology from the observed statistics of the Ly $\alpha$  forest and, conversely, predicting the properties of the Ly $\alpha$  forest from cosmological simulations (McQuinn, 2016, for a review).

## 1.2 Ly $\alpha$ Absorption in the IGM and the CGM

In the previous section, it has been explained that the Ly $\alpha$  forest arises from the scattering of photons along their path from the QSO to the observer. Solving the equations of radiative transfer for photons traveling from the background QSO to the observer through the IGM, one can obtain the expression of the Ly $\alpha$  opacity  $\tau_\nu$  of the IGM at position  $s$  and time  $t$ , as a function of the frequency  $\nu$  of the emitted photons in the rest

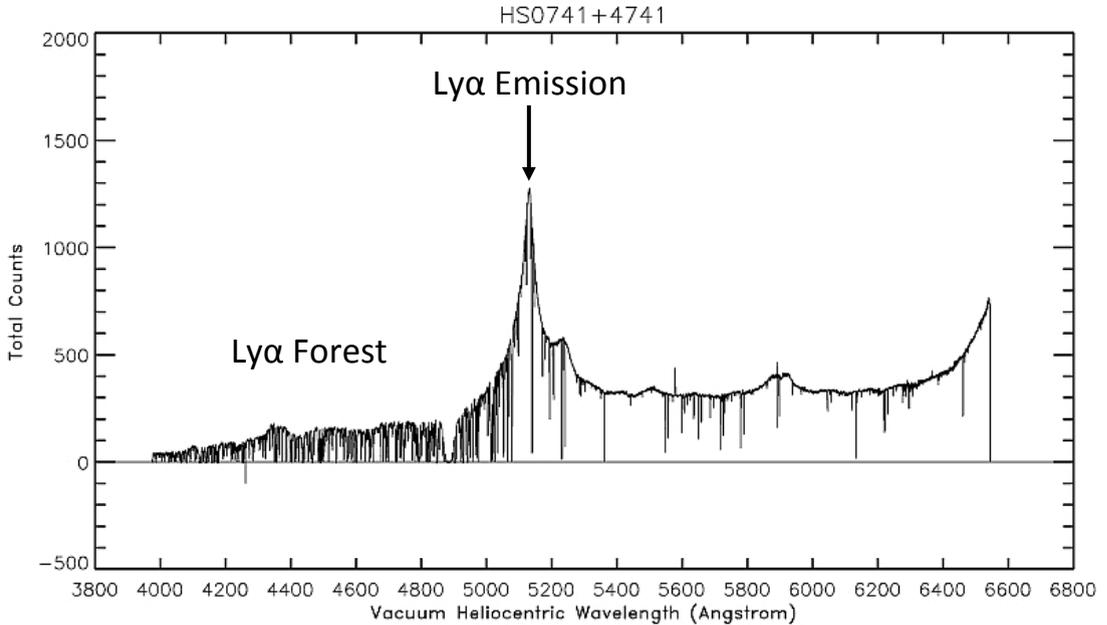


FIGURE 1.2: Spectrum of the quasar HS0741+4741 at  $z = 3.22$ . The peak at 5132 Å is the Ly $\alpha$  emission of the quasar. The Ly $\alpha$  forest is visible blueward of the peak. Redward of the peak, metal absorption lines can be seen. Figure adapted from Meiksin (2009), originally made by Songaila (2006).

frame of the QSO, located at position  $s_0$  (see, e.g., Meiksin, 2009):

$$\tau_\nu = \int_{s_0}^s ds' n_{\text{HI}}(s', t) \sigma_\nu, \quad (1.1)$$

where  $n_{\text{HI}}(s', t)$  is the HI number density and  $\sigma_\nu$  is the cross section of the scattering of Ly $\alpha$  photons off the line of sight. The flux, normalized to the continuum of the QSO, is then simply given by  $F = e^{-\tau}$ .

From (1.1), it is clear that the HI density plays a key role in determining the Ly $\alpha$  opacity. Therefore, when deriving the Ly $\alpha$  absorption in numerical simulations, it is important to accurately compute this quantity, modeling all relevant ionization and recombination mechanisms of hydrogen. Below the so-called self-shielding density threshold ( $\lesssim 6.0 \times 10^{-3} \text{ cm}^{-3}$  at  $2 < z < 3$ , see Rahmati et al. 2013), the ionized fraction of hydrogen is predominantly determined by the balance between photoionization due to the UVB and recombination, the rate of which is inversely proportional to the temperature of hydrogen ( $\propto T^{-0.7}$ ; Meiksin 2009). Above the self-shielding threshold, the photoionization rate drops down, because the hydrogen density is high enough so that the electrons that are stripped away from a newly photoionized hydrogen atom quickly

recombine with a neighboring HI nucleus. This is indeed the meaning of “self-shielding”, i.e. hydrogen “shields itself” from photoionization. On the other hand, in this density regime, collisions between hydrogen atoms are more frequent, and the electrons can be stripped away through collisional ionization. At even higher densities ( $\gtrsim 10^2 \text{ cm}^{-3}$ ), the photons emitted by newly recombined hydrogen atoms can, in turn, photoionize other hydrogen atoms (Rahmati et al., 2013). This recombination radiation contributes to the total photoionization rate, flattening its dependence on the hydrogen density in optically thick regions.

The typical densities of the IGM are well below the self-shielding threshold ( $\lesssim 10^{-4} \text{ cm}^{-3}$  at  $z \approx 2$ , see, e.g., Meiksin 2009), therefore the relevant physics of the Ly $\alpha$  forest is described by photoionization and recombination only. These processes can be easily implemented to compute the HI density from the baryon density distribution given by hydrodynamic cosmological simulations. The simplicity of the modeling makes simulations particularly effective in describing the properties of the IGM.

Within the CGM, the hydrogen density is often above the self-shielding threshold; as such, an accurate modeling has to take into account all the aforementioned ionization mechanisms. The most rigorous way to model them would be running full radiative transfer simulations (e.g., Faucher-Giguère et al., 2009a, Iliev et al., 2007, Petkova & Springel, 2011, Zahn et al., 2011), but this approach is computationally very expensive. For this reason, the various ionization mechanisms are usually implemented with analytical approximations, previously calibrated through radiative transfer codes, on top of the outputs given by hydrodynamic simulations (e.g., Rahmati et al., 2013). Whereas the physics governing Ly $\alpha$  absorption is straightforward and well understood, the density and temperature of the CGM depend also on less clear processes, for instance outflows, gas accretion and galactic winds. These processes are generally implemented through active galactic nuclei (AGN) and stellar feedback prescriptions in cosmological simulations. This means that the Ly $\alpha$  absorption in the CGM could potentially constrain feedback and other sub-resolution physics in simulations.

A full account of all the literature about observations and simulations of Ly $\alpha$  absorption in the IGM and the CGM is beyond the scope of this Chapter. The introductions of Chapters 2 and 3 include a review of the numerical works studying Ly $\alpha$  absorption in the IGM and the CGM that are relevant to the research projects described in those chapters.

### 1.3 Thesis Outline

This thesis presents the results of my work undertaken mainly at the Max-Planck-Institute for Astronomy (Max-Planck-Institut für Astronomie), under the supervision of Prof. Dr. Joseph F. Hennawi. The main purpose of the thesis is to exploit the full

potential of the Ly $\alpha$  absorption line to investigate the physics of the IGM and the CGM through cosmological simulations.

The first part of the thesis focuses on the IGM. Chapter 2 presents the result of the my first PhD project, published in Sorini et al. (2016). I develop a novel technique to predict various statistics of the Ly $\alpha$  forest with large N-body cosmological simulations. I demonstrate that the results given by this technique agree within 10%-13% (depending on the specific statistic considered) with the predictions of a reference hydrodynamic simulation. By comparing the results of the method with other attempts of modeling the statistics of the Ly $\alpha$  forest in the literature, I display that the technique developed represents the state-of-the art in this respect. I also show that, given current numerical constraints, the method developed in my work could be applied on Gpc-scale N-body simulations. Therefore, its possible applications include a more accurate estimation of the scale of the Baryon Acoustic Oscillations (BAO) from their signature on the cross-correlation function of the Ly $\alpha$  forest, or the study of large-scale fluctuations of the UVB.

The second part of the thesis considers both the IGM and the CGM. In Chapter 3, I present the second project undertaken during my PhD (Sorini et al., 2017, to be submitted). For the first time, I compare the predictions of state-of-the-art hydrodynamic cosmological simulations with observations of Ly $\alpha$  absorption around foreground galaxies at different transverse separations from background quasars, ranging from  $\sim 25$  kpc to  $\sim 17$  Mpc. The main result is that the observables that I considered are capable of constraining the physics implemented in numerical simulations (such as feedback prescriptions), even more so with the increasingly high precision expected from surveys in the near future. Thus, the comparison presented in Chapter 3 represents a new fundamental test for simulations that should be considered by future numerical works.

Chapter 4 contains the preliminary results that I obtained within a project under the supervision of Prof. Dr. Matthias Bartelmann, at the Institute for Theoretical Astrophysics (Institut für Theoretische Astrophysik). Although, as underlined in the first chapters, cosmological simulations represent the most efficient way to study Ly $\alpha$  absorption and in general the large-scale structure of the Universe, a major need still remains for developing analytical methods capable of predicting statistics of the clustering of matter. In this context, I present numerical tests that will be needed in the near future, in order to compare the power spectrum of the momentum density fluctuations of DM, computed through a novel analytic method based on Bartelmann et al. (2016), with the predictions of N-body simulations (Littek, Sorini et al., in prep.). This represents an intermediate step necessary to generalize Bartelmann et al. (2016) approach in order to achieve the first accurate, fully analytical, prediction of the statistics of Ly $\alpha$  forest statistics (Sorini et al., in prep.).

I summarize the main conclusions of this thesis in Chapter 5, where I also discuss the perspectives opened by my research. I am the main contributor of the work presented

in this thesis. At the beginning of every chapter, I will clearly state what my personal contribution is.

During my PhD studies, I have also independently accomplished an extensive follow-up work based on my Master thesis, which resulted in a publication (Sorini, 2017), of which I am the sole author. This work is not included in this thesis, because it focuses on slightly different topics and was partially already included in my Master thesis. My list of publications can be found at the end of this thesis, before the bibliography.



## Chapter 2

# Modeling the Ly $\alpha$ Forest in Collisionless Simulations

The large number of QSOs discovered to date enables statistical analyses of the absorption spectra by considering the Ly $\alpha$  transmitted flux along many different lines of sight, called “skewers”. The measured statistical properties can be compared to theoretical models of the IGM, constraining cosmological parameters as well as the thermal history of the IGM. In this work, we focus on three observationally most relevant statistics of the transmitted flux: the probability density function (PDF; Rauch et al. 1997) the line-of-sight power spectrum (1DPS; Croft et al. 1999, Palanque-Delabrouille et al. 2015), and the 3D power spectrum (3DPS; Slosar et al. 2011). The 3DPS can be used for an independent measurement of the BAO characteristic scale (Delubac et al., 2015, Font-Ribera et al., 2014a); future increase in the number of observed quasars at redshifts  $z > 2$  promises tight constraints on the expansion history of the universe at high redshifts and other cosmological parameters (Font-Ribera et al., 2014b).

---

This Chapter contains the material published in Sorini et al. (2016), slightly re-adapted for this thesis. I am the main contributor of the work described here. I developed several ideas underlying the methods presented in this Chapter and implemented most of the codes utilized for the analysis of the simulations, which was conducted entirely by myself. Dr. José Oñorbe contributed with ideas and advice on the development of the analysis and the codes. Dr. Zarija Lukić ran the simulations considered in this work and provided suggestions for tests to be conducted. Prof. Dr. Joseph F. Hennawi is the author of some of the codes used in the analysis of the simulations. He owns the conceptual foundation of the work and contributed to the project with ideas and advice.

When this work was in its early stage, highly preliminary results were presented in the dissertation for the Galilean Diploma, submitted to the Galilean School of Higher Education (Scuola Galileiana di Studi Superiori) of Padua (Italy) in November 2014. Those results concern numeric tests functional to the later development of methods discussed in this Chapter. Such tests were based on simulations at a lower resolution than the ones used in this thesis. Hence, despite the similar topic, the results and simulations presented in this Chapter are clearly different from the ones presented in the earlier Galilean dissertation (which is unpublished). As such, the results presented in the Galilean dissertation do not constitute part of this thesis.

Performing all the above mentioned studies requires not only precise observations, but also accurate theoretical modeling that is far from being straightforward. As introduced in Chapter 1, the Ly $\alpha$  forest is the observational signature of HI, and is set by the interplay of gravitational collapse, expansion of the universe, and reionization processes due to the buildup of a background of UV photons emitted by AGN and star-forming galaxies (Cen et al., 1994, Croft et al., 2002, Hernquist et al., 1996, McDonald et al., 2000, Meiksin & White, 2001, Zhang et al., 1997). There is no analytic solution for the small-scale evolution of the (baryon) density fluctuations over time. In order to precisely describe the behavior of the IGM, it is therefore necessary to treat the problem numerically. In this respect, hydrodynamic cosmological simulations have led to a consistent description of the IGM in the framework of structure formation (Cen et al., 1994). However, they are computationally expensive, making it challenging to reach high resolutions. Furthermore, available memory limits how large volume can be run in high-resolution simulations. For example, it would be necessary to run a simulation of  $\sim 1$  cGpc on a side to probe the scales of BAO and study their signature in the Ly $\alpha$  forest (Norman et al., 2009, Slosar et al., 2009, White et al., 2010). The absorption lines are set by physical processes occurring around the Jeans scale (see, e.g., Dodelson, 2003), whose order of magnitude is expected to be 100 ckpc (Gnedin & Hui, 1996, 1998, Kulkarni et al., 2015, Rorai et al., 2013). Recent work indicates that a resolution of 20 ckpc is required to achieve  $\sim 1\%$  precision in the description of the statistics of the Ly $\alpha$  forest (Lukić et al., 2015). This implies that IGM-BAO simulations would require at least  $50000^3$  resolution elements to span such a wide dynamic range, far beyond current (and near future) computational resources.

Collisionless simulations neglect baryonic pressure, therefore they are not as accurate as hydrodynamic simulations on small scales. However, on large scales baryonic forces are negligible, thus collisionless simulations are as good as hydrodynamic ones in this regime. For this reason, N-body collisionless simulations are often used in cosmology to study the formation and evolution of structure in large volumes, but with poor mean inter-particle spacing (often several hundreds ckpc). Clearly, it is desirable to find strategies that combine the volume of collisionless but retain the accuracy of high-resolution hydrodynamic simulations. This objective has been recognized in the past, resulting in the development of various approximate methods to predict the Ly $\alpha$  forest from N-body simulations.

The simplest approach is assuming that baryons perfectly trace DM (e.g. Croft et al. 1998, Petitjean et al. 1995). In this over-simplified picture, the baryon density field is the scaled version of the matter density field. However, DM particles are collisionless, so the pressure of baryons which competes with gravitational collapse is simply neglected. The effect of pressure was instead included by e.g. Gnedin & Hui (1998) as a modification of the gravitational potential. Another approach consists in running three simulations, all with insufficient resolution or box size, and compensate for the resulting errors in the estimation of the power spectrum applying splicing techniques (Borde et al., 2014,

McDonald, 2003). It has been shown by Lukić et al. (2015) that the accuracy of this method is around 10%. A different widely used strategy is mimicking baryon pressure by smoothing the matter density field with a Gaussian kernel (Gnedin & Hui, 1996, Meiksin & White, 2001, Rorai et al., 2013, Viel et al., 2006, 2002). The flux field produced by the smoothed density field can then be computed imposing a polytropic temperature-density relationship to the IGM (Hui & Gnedin, 1997). This method reproduces the statistics of the flux field reasonably well. For example, Meiksin & White (2001) claim 10% agreement between Gaussian-smoothed collisionless simulations and hydrodynamic simulations in the cumulative distribution of the flux.

A more refined way to reconstruct the baryon density is applying *ad hoc* transformations to the matter density field, calibrated with a reference hydrodynamic simulation (Viel et al., 2002). Mocks of Ly $\alpha$  forest spectra can be obtained generating a Gaussian random field and then transforming it so that it matches a certain flux PDF and power spectrum (Bautista et al., 2015, Font-Ribera et al., 2012a, Font-Ribera & Miralda-Escudé, 2012). Recently, Peirani et al. (2014) exploited a hydrodynamic simulation to calibrate a mapping from the density field of an N-body simulation to the Ly $\alpha$  forest flux, tuned to reproduce the PDF of the flux. Then, artificial flux skewers are created in order to reproduce the two-point function of the flux given by the calibrating simulation. This is done first by computing the conditional PDF of the flux, given the DM density, from the reference simulation. Subsequently, each pixel is assigned a value of such “conditional flux”. This procedure seems to yield reasonable correlation functions, but noisy skewers as well. This problem is remedied by drawing flux values from the Gaussianized percentile distribution of the conditional flux, and then forcing it to match the PDF of the conditional flux. Visually examining the plots of the resulting flux power spectrum, it appears close to the one provided by the reference hydrodynamic simulation, but the accuracy is not quantified by the authors.

The lack of quantitative assessments in the literature makes it harder to compare the results obtained via different methods. Conducting a more quantitative study is important for establishing which problems can be addressed by what methods. Another important point regards the value of the filtering scale generally adopted in the Gaussian smoothing of matter. The value of the filtering scale has been measured from observational data only recently (70 – 120 kpc in the redshift range  $2 < z < 3.6$ , see Rorai et al. 2017)<sup>1</sup>. Hence, in previous numerical studies it has been set to “reasonable” values, in any case not smaller than the mean interparticle spacing of the simulations involved (otherwise the smoothing would have negligible effect). For example, White et al. (2010) used 139 ckpc as smoothing scale in a simulation with a box size of 1.02 cGpc and 4000<sup>3</sup> particles. Other authors have chosen larger values, for example Peirani et al. (2014) compare their method with Gaussian-smoothed DM simulations with a filtering scale of  $300 h^{-1}$  ckpc and  $1 h^{-1}$  cMpc.

<sup>1</sup>This was not known at the time the work presented in this Chapter was carried out.

In the work presented in this Chapter, we use the Gaussian smoothing technique as a starting point upon which we add more refined transformations of the matter density field. Following this line of reasoning, we develop two methods, named 1D-IMS and 3D-IMS, where IMS stands for “Iteratively Matched Statistics”, the technique on which they are grounded. The purpose of our methods is to accurately obtain the flux statistics from collisionless simulations. This is done through hydro-calibrated mappings, which are conceptually simpler than those adopted by Peirani et al. (2014). We quantify how accurately our methods reproduce the PDF, 1DPS and 3DPS of the flux given by a reference hydrodynamic simulation.

The high accuracy of our methods and a weak dependence on the initial smoothing scale represent a clear advantage over the Gaussian smoothing technique. Our methods thus enable using large-box collisionless simulations which do not resolve the Jeans scale. One important application we have in mind is modeling the BAO signature in the Ly $\alpha$  forest. However, there are more topics which can benefit from it: studies of UVB fluctuations, cross-correlations between galaxies and Ly $\alpha$  forest and others.

This Chapter is organized as follows. In § 2.1 we describe our simulations and calculation of Ly $\alpha$  flux. In § 2.2 we discuss the impact of the most important assumptions underlying approximate techniques to predict the Ly $\alpha$  forest in collisionless simulations. The Gaussian smoothing method is explored into great detail and we present the first quantitative analysis of its accuracy in reproducing the 3DPS of flux, as a function of the smoothing length. In § 2.3 we describe 3D-IMS and 1D-IMS, assessing their accuracy. We compare the performances of the various methods in § 2.4. In § 2.5, we apply 3D-IMS to a future relevant context: we compute the flux statistics through an N-body simulation, calibrating the transformations involved in our technique with a smaller hydrodynamic simulation. We show that the method retains accuracy, while at the same time we demonstrate that the Gaussian smoothing technique does not yield accurate predictions when applied to simulations involving large boxes. We also compare the techniques considered by us with previous work in § 2.6. Finally, in § 2.7 we present the conclusions of this Chapter, discussing possible future applications of our work as well.

## 2.1 Simulations

The hydrodynamic simulations that we use in this Chapter are carried out with Nyx code (Almgren et al., 2013, Lukić et al., 2015), while N-body runs are performed with Gadget code (Springel, 2005). Both codes employ leapfrog — second order accurate method for integrating the equations of motion of particles. Both codes also adopt the particle-mesh (PM) method with cloud-in-cell (CIC) interpolation for calculating gravitational forces. On top of the PM calculation, Gadget adds gravitational short-range force using Barnes-Hut (Barnes & Hut, 1986) hierarchical tree algorithm, therefore going to the higher resolution than our Nyx runs done on a uniform Cartesian grid. Nyx, in addition to

gravity, solves equations of gas dynamics using second-order accurate piecewise parabolic method. To better reproduce the 3D fluid flow, a dimensionally unsplit scheme with full corner coupling is adopted (Colella, 1990). Heating and cooling are integrated using VODE (Brown et al., 1989) and are coupled to hydrodynamics through Strang splitting (Strang, 1968). All cells are assumed to be optically thin and radiative feedback is considered only through the UVB model given by Haardt & Madau (2012). For cooling rates and further details on the physics in Nyx simulations, we refer the reader to Lukić et al. (2015) paper. The cosmological model assumed is the  $\Lambda$ CDM model with parameters consistent with the 7-year data release of WMAP<sup>2</sup> (Komatsu et al., 2011):  $\Omega_m = 0.275$ ,  $\Omega_\Lambda = 1 - \Omega_m = 0.725$ ,  $\Omega_b = 0.046$ ,  $h = 0.702$ ,  $\sigma_8 = 0.816$ ,  $n_s = 0.96$ . Six baryonic species are implemented:  $e^-$ , HI, HII, HeI, HeII, HeIII. The simulations are initialized at  $z = 159$  with a grid distribution of particles and Zel’dovich approximation (Zel’dovich, 1970).

To recover the absorption spectra from our simulations, we choose the lines of sight, which we refer to as “skewers”, drawn parallel to one of the sides of the simulation box. The optical depth is computed according to equation (2.1). After extracting skewers, we rescale the optical depth so that the mean flux is  $\langle F \rangle = 0.68$  at  $z = 3$ , a value consistent with current observations (Becker & Bolton, 2013, Faucher-Giguère et al., 2008a). Unless otherwise indicated, the results presented in this work refer to redshift  $z = 3$ , but in order to confirm our conclusions are not dependent on redshift, we have also analyzed redshifts  $z = 2$  and  $z = 4$ .

In this work we use two hydrodynamic and two N-body simulations. The two Nyx hydrodynamic simulations have identical physics and the same spatial resolution, differing only in the choice of the box size: the smaller one has a box of 14.2 cMpc on a side, while the larger one 114 cMpc. The two simulations have  $512^3$  and  $4096^3$  resolution elements respectively, and they were a part of the convergence study done in Lukić et al. (2015). We will first use only one Nyx simulation to test how well we can reproduce the forest statistics given only DM particles and no gas information. This simulation is  $\sim 1\%$  converged resolution-wise, but the box size is too small for accurate reproduction of flux statistics. However, the main point of our work is to test how well we can match the given flux statistics, and it is irrelevant how accurately that statistics is describing a particular cosmological model. In other words, it is important to have a resolution good enough to correctly capture small-scale physics, but it does not matter that the large-scale power is missing in the simulation.

We then want to test how accurately the forest statistics can be reproduced in large-volume simulations, i.e. with box sizes of 1 cGpc and larger. Of course, we do not have the “true” answer for such large boxes as it would be obtained with hydrodynamic simulations. Instead, we will use the results of the large-box Nyx run, which is demonstrated to be converged both in resolution and box size (Lukić et al., 2015), as the

---

<sup>2</sup>Wilkinson Microwave Anisotropy Probe

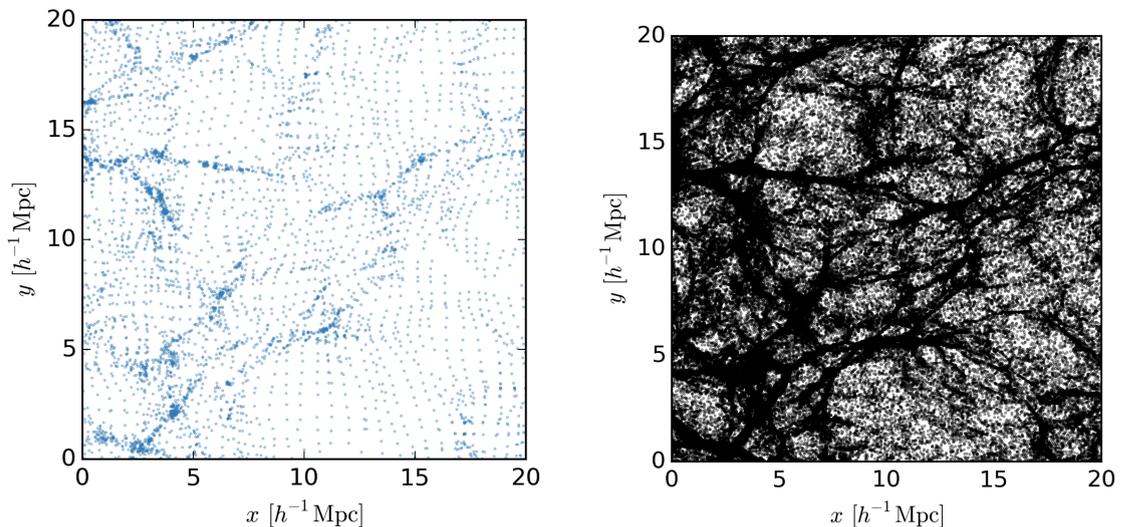


FIGURE 2.1: Difference in mass resolution between a state-of-the-art “Hubble-volume” simulation and a hydrodynamic simulation targeting the Ly $\alpha$  forest. On the left (blue points) we show all particles in  $0.5 h^{-1} \text{Mpc}$  thick slice from the  $256^3$  Gadget N-body run. That corresponds to a trillion particles in a  $4 \text{cGpc}$  box simulation. Right panel (black points) shows only 1% of particles in the same region from the  $4096^3$  Nyx hydrodynamic run, visually demonstrating the level of detail needed to capture flux statistics at percent-level accuracy.

“truth”, and we will reconstruct its flux statistics using an N-body Gadget run in same box together through the small-box Nyx run. We ran two Gadget N-body simulations with the same box size as the larger Nyx simulation,  $114 \text{cMpc}$ , but with only  $512^3$  and  $256^3$  particles. The number of particles in Gadget runs is chosen to be representative of the mean inter-particle spacing in the state-of-the-art N-body simulations of “Hubble” volumes (e.g Habib et al. 2012, 2013, Skillman et al. 2014). As an example, we show in the left panel of Figure 2.1 all particles in a  $0.7 \text{cMpc}$  thick slice from the  $256^3$  Gadget run. This run in  $114 \text{cMpc}$  box yields approximately the same mass resolution as one trillion particles in a  $4 \text{cGpc}$  simulation would. The right panel displays only 1% of the particles in the same region, from the  $4096^3$  Nyx hydrodynamic run. The comparison of the two panels makes it immediately apparent the high resolution which is needed for modeling the Ly $\alpha$  forest statistics at  $\sim 1\%$  accuracy.

Gadget simulations share the same phases in the initial conditions as the large Nyx run (as clearly visible in Figure 2.1), enabling comparison of individual skewers. We emphasize that although somewhat artificial, this test is actually more difficult than the real-world situation, and thus we expect that we can only overestimate the error of our method. The reason is that in reality we would use a fully converged  $\sim 142 \text{cMpc}$  hydrodynamic simulation to model flux statistics in  $\sim 1 \text{cGpc}$  N-body simulations, making box size errors negligible, whereas here we cannot avoid them.

### 2.1.1 Ly $\alpha$ Skewers

The Ly $\alpha$  forest arises from the scattering of photons along their path from a background quasar to the observer. The fraction of the transmitted flux is  $F = \exp(-\tau)$ , where  $\tau$  is the opacity of the intervening IGM. The opacity in redshift space at a given velocity coordinate  $u$  along the line of sight is given by

$$\tau(u) = \int du' \frac{\lambda_{\text{Ly}\alpha} \sigma n_{\text{HI}}(\mathbf{u}')}{H(z)b(\mathbf{u}')} \exp \left[ -\frac{(u - u_0(\mathbf{u}'))^2}{b(\mathbf{u}')^2} \right] \quad (2.1)$$

where  $u'$  is the component of the Hubble flow velocity field  $\mathbf{u}'$  along the line-of-sight, over which the integral is calculated. In the above expression,  $n_{\text{HI}}(\mathbf{u}')$  is the number density of neutral hydrogen and  $\sigma$  and  $\lambda_{\text{Ly}\alpha}$  are the cross section<sup>3</sup> and wavelength of the Ly $\alpha$  transition in the rest frame, respectively. The line-of-sight velocity of gas particles is then given by  $u_0(\mathbf{u}') = u' + u_{\text{pec}}(\mathbf{u}')$ , the second term being the peculiar velocity of the gas. In equation (2.1) thermal broadening is described by  $b(\mathbf{u}') = \sqrt{2k_B T(\mathbf{u}')/m_p}$ , where  $T(\mathbf{u}')$  is the temperature of the gas and  $m_p$  the proton mass. The convolution with thermal broadening and peculiar velocities actually yields a Voigt profile in equation (2.1) instead of a Gaussian. However, the latter is a good approximation for  $\tau < 100$  (Lukić et al., 2015), regime relevant for the Ly $\alpha$  forest studies. Computation of  $\tau$  requires a determination of the neutral hydrogen density, which in turn depends on baryon density and temperature, as well as the hydrogen ionization and recombination rates. The challenge for approximate methods is to recover relevant Ly $\alpha$  forest statistics, without the knowledge of baryon thermodynamical quantities.

## 2.2 Limitations of Approximate Methods

The very first task of approximate methods is to obtain an estimate for the baryon density field. This is commonly done via manipulation of the density field in an N-body run (Meiksin & White, 2001, Peirani et al., 2014, Viel et al., 2002), to account for the baryonic pressure smoothing. The functional form for the smoothing is usually a Gaussian and that is indeed the starting point for all methods considered in this work. Secondly, approximate methods need other assumptions, concerning the estimate of the temperature of the IGM and its velocity field. In this section, we review these approximations and assess their impact on the accuracy of the Ly $\alpha$  flux, as a function of the smoothing length. In order to better understand inherent limitations of the approximate methods, we will also consider separately the accuracy of baryon density reconstruction of other thermodynamic quantities.

<sup>3</sup>Actually, when one considers thermal motions of the gas particles, the cross section  $\sigma_{\text{Ly}\alpha}$  of the Ly $\alpha$  transition is given by the product of  $\sigma$  and a Voigt profile. Integrating  $\sigma_{\text{Ly}\alpha}$  over all possible frequencies of the intervening photon, one obtains  $\sigma$ . Therefore, strictly speaking,  $\sigma$  is the frequency-integrated cross section and, as such, has the dimensions of area/time. For an extensive derivation of equation (2.1), see e.g. Meiksin (2009).

### 2.2.1 Gaussian Smoothing

A pseudo baryon density field can be generated from a collisionless simulation by smoothing the matter density fluctuations  $\delta$  at a characteristic smoothing length  $\lambda_G$  given as

$$\delta^{\lambda_G}(\mathbf{k}) = \delta(\mathbf{k}) \exp(-\lambda_G^2 k^2). \quad (2.2)$$

This length is expected to be of the order of the Jeans filtering scale, which is in comoving units (Binney & Tremaine, 2008):

$$\lambda_J^2(t) = \frac{c_s^2(t)a(t)}{4\pi G\rho_0}, \quad (2.3)$$

where  $c_s$  is the speed of sound at time  $t$ ,  $a(t)$  the scale factor and  $\rho_0$  the mean matter density and  $G$  is Newton's gravitation constant. The same line of reasoning can be applied to the line-of-sight velocities of particles as well. Once both matter density and velocities are smoothed, flux skewers can be computed with some approximation for the IGM temperature (discussed in § 2.2.2), replacing baryon density and velocity fields with the corresponding smoothed matter quantities. For the sake of clarity, we summarize the inputs required to apply the Gaussian smoothing technique (and our methods, which will be discussed in § 2.3.1 and § 2.3.2) in Table 2.1.

There are quantitative studies in the literature aiming to understand how well the Gaussian smoothing technique reproduces various flux statistics computed through hydrodynamic simulations (see § 2.6 for details), but none of them considers the flux 3DPS. We take  $\lambda_G$  as a free parameter and assess the accuracy with which the Gaussian smoothing technique recovers the flux 1DPS, PDF and 3DPS, through the following steps:

1. We have particle positions and velocities from a simulation. We deposit them on a grid using CIC deposition; we use the grid with as many cells as the number of particles in the simulation.
2. We smooth this density field with a certain smoothing scale  $\lambda_G$  and the velocity field at 228 ckpc (see appendix A).
3. We compute 1DPS, 3DPS and PDF of the flux field obtained using Fluctuating Gunn-Peterson Approximation (see § 2.2.2).

### 2.2.2 Fluctuating Gunn-Peterson Approximation

Equation (2.1) can be simplified expressing the neutral hydrogen density  $n_{\text{HI}}$  as a function of the baryon density fluctuations  $\delta_b$ . Let us consider a gas composed by hydrogen and helium. Let  $x_{\text{HII}}$ ,  $x_{\text{HeII}}$  and  $x_{\text{HeIII}}$  be the fractions of ionized hydrogen, singly and

doubly ionized helium respectively. The total number densities of hydrogen  $n_{\text{H}}$  and helium  $n_{\text{He}}$  are related through  $n_{\text{He}} = \chi n_{\text{H}}$ , where  $\chi = X/4Y$ . Assuming photoionization equilibrium, the number density of neutral hydrogen is given by

$$n_{\text{HI}} = \frac{\alpha(T)}{\Gamma_{\text{HI}}} x_{\text{HII}} [(1 + \chi)x_{\text{HII}} + \chi x_{\text{HeIII}}] n_{\text{H}}^2 \quad (2.4)$$

where  $\alpha(T) \propto T^{-0.7}$  is the Case A recombination coefficient per proton and  $\Gamma_{\text{HI}}$  the photoionization rate of hydrogen. Commonly used Case A and B definitions differentiate media that allow the Lyman photons to escape or that are opaque to these lines (except for Lyman-alpha), respectively. Case A is used in the optically-thin limit, which can be applied in most of the Ly $\alpha$  forest, since it describes regions where the density is low enough to let Lyman limit photons escape (Furlanetto et al. 2006, Kuhlen & Faucher-Giguère 2012; see also the discussion in Miralda-Escudé 2003 and Kaurov & Gnedin 2014). If helium is only singly ionized, the factor between square brackets in equation (2.4) becomes  $(1 + \chi)x_{\text{HII}}$ , while for  $x_{\text{HII}} = x_{\text{HeIII}}$  it is  $(1 + 2\chi)x_{\text{HII}}$ . Apart from the detailed modeling of the ionized fractions, the important point of equation (2.4) in this context is that  $n_{\text{HI}} \propto T^{-0.7} n_{\text{H}}^2$ .

Simulations show that the temperature-density relationship of the IGM is a power law over a wide range of density and temperature (Hui & Gnedin, 1997), so that

$$T(\mathbf{u}) = T_0 (1 + \delta_b(\mathbf{u}))^{\gamma-1} \quad (2.5)$$

where  $T_0$  and  $\gamma$  are constants. From our simulation, at redshift  $z = 3$ , we obtained  $T_0 = 1.09 \times 10^4 \text{ K}$  and  $\gamma = 1.56$ , following the fitting procedure described by Lukić et al. (2015). Assuming (2.5), the relationship between  $n_{\text{HI}}$  and  $\delta_b$  can be expressed in terms of the parameters of our simulation as follows:

$$n_{\text{HI}}(\mathbf{u}) = A \frac{8.28 \times 10^{-13} \text{ s}^{-1}}{\Gamma_{\text{HI}}} \frac{\Omega_b h^2}{0.0227} \left( \frac{1+z}{4} \right)^3 \left( \frac{T_0}{1.09 \times 10^4 \text{ K}} \right)^{-0.7} [1 + \delta_b(\mathbf{u})]^{2-0.7(\gamma-1)} \quad (2.6)$$

where  $A$  is a proportionality constant. For our simulation,  $A = 3.09 \times 10^{-12} \text{ cm}^{-3}$ . Neglecting the scatter in the temperature-density relationship of the IGM, i.e. assuming (2.5) and consequently  $n_{\text{HI}} \propto (1 + \delta_b)^{2-0.7(\gamma-1)}$ , is usually referred to as ‘‘Fluctuating Gunn-Peterson Approximation’’ (FGPA; Croft et al. 1998, Weinberg et al. 1997). Since the FGPA is useful when one cannot or does not wish to run a hydrodynamic simulation, one also needs an approximation for  $\delta_b$  in equation (2.6). For this reason, in any practical situation  $\delta_b$  is replaced by the DM density fluctuations  $\delta_{\text{DM}}$ , with or even without Gaussian smoothing. For the sake of clarity, in the remainder of our work we shall refer solely to the operation described by equation (2.2) with ‘‘Gaussian smoothing’’. On the contrary, the Gaussian smoothing of the DM density field, combined with the FGPA

to compute the Ly $\alpha$  flux field, shall be denoted as ‘‘Gaussian smoothing and FGPA’’ (GS+FGPA).

We now define a new field, the ‘‘flux in real space’’ (or simply ‘‘real flux’’)  $F_{\text{real}}$  as the flux that would be obtained neglecting thermal broadening and peculiar velocities. This is not a physical observable, but the shape of its power spectrum is sensitive to the Jeans scale (Kulkarni et al., 2015) and it will be a useful quantity in our computations. As such, we can define the opacity in real space

$$\tau_{\text{real}}(\mathbf{u}) = \frac{\lambda_{\text{Ly}\alpha}\sigma}{H(z)} n_{\text{HI}}(\mathbf{u}) \quad (2.7)$$

Within the FGPA,  $\tau_{\text{real}} \propto n_{\text{HI}} \propto (1 + \delta_b)^{2-0.7(\gamma-1)}$ . Convolution (2.7) with the gas velocities and thermal broadening, one obtains (2.1).

### 2.2.3 Accuracy of FGPA

We now assess the accuracy of the FGPA using the baryon density field from the hydrodynamic simulation as a reference. Indeed, we want to focus on how the accuracy of the FGPA is influenced by approximating the baryon velocity field with the Gaussian-smoothed DM velocity field and by the assumption that the temperature-density relationship is a power-law. In this way, we investigate the ‘‘inherent’’ accuracy of the FGPA.

In our hydro simulation, we also have the velocities of DM particles. Thus we construct the velocity field by CIC-binning them on a grid with as many cells as the number of particles and then smooth it with a Gaussian kernel. In principle, the smoothing length of velocity could be different from the one of the DM density field. We keep it fixed to 228 ckpc throughout this work, since we verified that this value gives the best overall accuracy in reproducing the statistics considered (see appendix A for further details). However, we have also checked that modifying the smoothing length for the velocity field does not significantly change our conclusions.

In Figure 2.2 we show different physical quantities along one skewer as an example, to display the differences between the hydrodynamic simulation (solid green lines) and the FGPA (dashed blue lines). The top panel shows the density fluctuations along the skewer considered. The second panel underscores the differences between a temperature-density relationship with no scatter and the temperature given by the hydrodynamic simulation. We see that the biggest differences arise around the highest density peaks, where shocks could be present. In the third panel we plot the line-of-sight velocity of DM particles<sup>4</sup> (black line) and baryons (green line). Here we also plot the smoothed DM velocity, which

<sup>4</sup>The CIC-binned velocity field of DM particles occasionally results in pixels with no particles in them. To correct for this effect, we assign to these grid cells the average velocity of their first neighbors in the 3D space. Then, we proceed with the Gaussian smoothing.

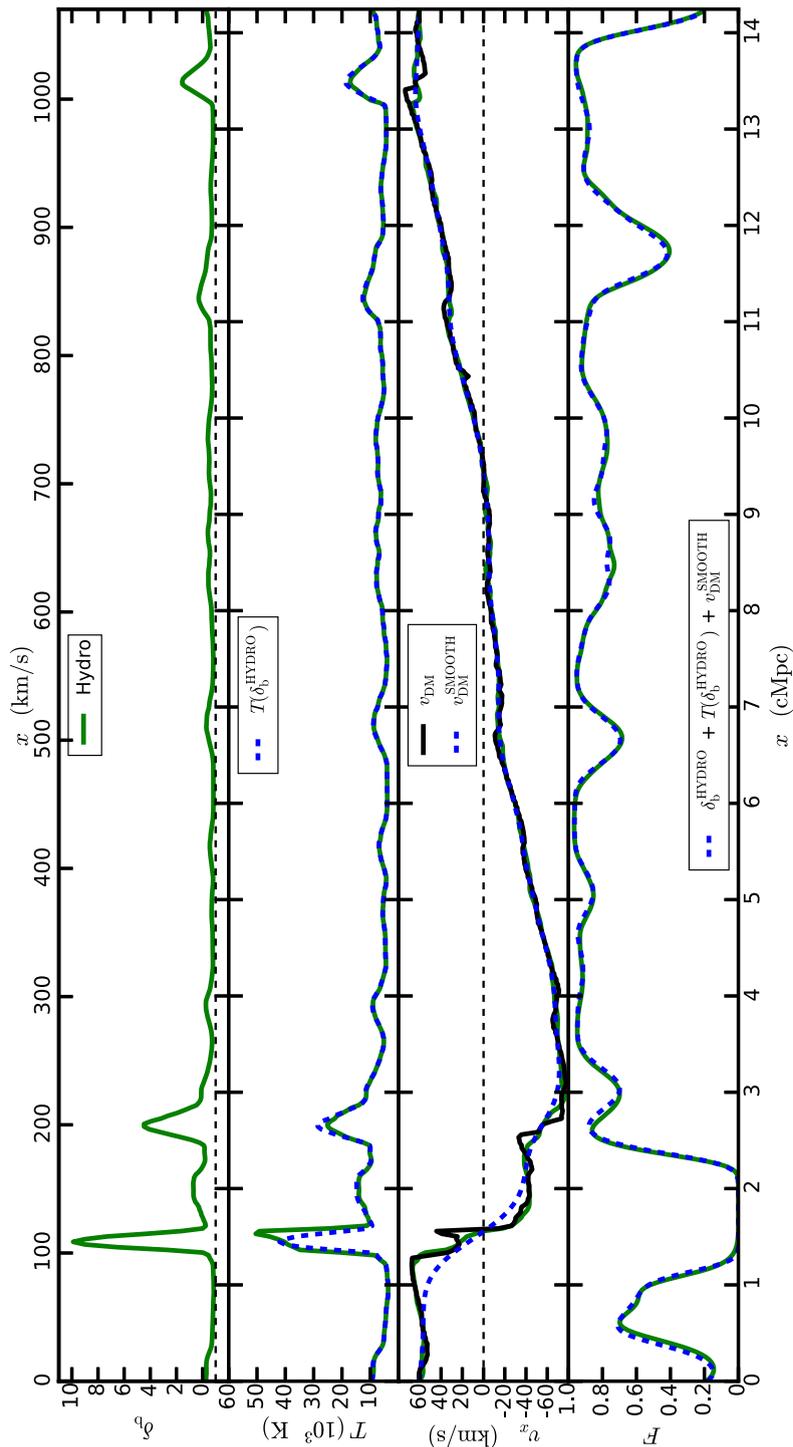


FIGURE 2-2: Different quantities along a certain skewer are plotted, to illustrate possible limitations of the FGPA. *First panel:* Baryon density fluctuations from the hydrodynamic simulation. *Second panel:* Temperature obtained from the hydrodynamic simulation (solid green line) and by imposing a 1-to-1 temperature-density relationship (see text for details) to the baryon density given by the simulation (dashed blue line). *Third panel:* Line-of-sight velocity of baryons (green line) and dark matter (black line), obtained directly from the hydrodynamic simulation. The dashed blue line represents the line-of-sight velocity obtained smoothing the DM velocity with a smoothing scale of 228 cMpc. *Fourth panel:* Flux obtained from the hydrodynamic simulation (solid green line) and the one obtained by imposing a deterministic temperature-density relationship to the baryon density given by the simulation, and using the Gaussian-smoothed line-of-sight velocities of dark matter instead of baryons (dashed blue line).

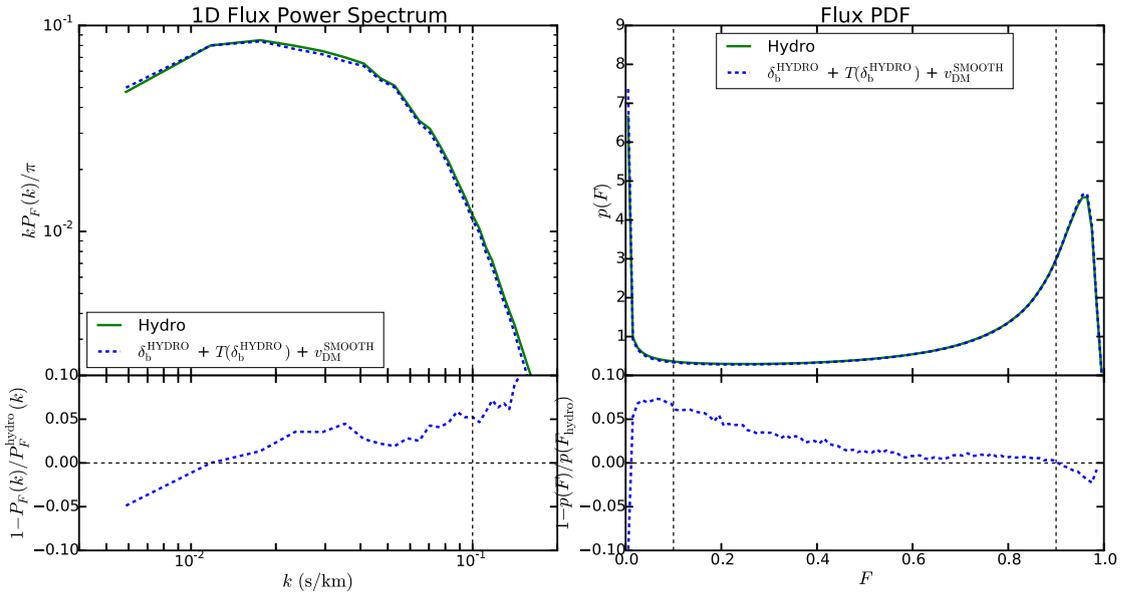


FIGURE 2.3: In the top panels, the solid green lines represent the dimensionless 1DPS (left) and PDF (right) of the flux given by our reference hydrodynamic simulation. The dashed blue lines are the 1DPS and PDF of the flux computed by imposing a 1-to-1 temperature-density relationship on the baryon density given by the hydrodynamic simulation, and using the Gaussian-smoothed line-of-sight velocities of dark matter instead of baryons. The dashed vertical line delimits the dynamic range considered to compute the accuracy (see text for details). The relative errors plotted in the lower panels set the intrinsic limitations of approximate techniques predicting the Ly $\alpha$  forest through the manipulation of the DM density field given by collisionless simulations.

is the one we actually adopt (dashed blue line). In the last panel, we show the difference between the flux computed as explained in this section and from the hydrodynamic simulation. We notice that the FGPA recovers the flux skewer remarkably well.

We show the results about the statistics of flux skewers in Figures 2.3 and 2.4. In the upper panels of Figure 2.3 we show the flux 1DPS and PDF given by the hydrodynamic simulation and the FGPA applied as explained above. In the lower panels, we show the relative difference of the statistics obtained with respect to the results of the reference simulation. Analogous plots for the flux 3DPS can be seen in Figure 2.4. The 3DPS can be expressed as a function of the norm of the  $\mathbf{k}$ -mode considered and of  $\mu = \hat{\mathbf{n}} \cdot \mathbf{k}/k$ , where  $\hat{\mathbf{n}}$  is the unit vector parallel to the line-of-sight. We shall denote the dimensionless 3DPS as  $\Delta^2(k, \mu) = k^3 P_F(k, \mu)/2\pi^2$ .

The accuracy of the FGPA of course depends on the Fourier modes considered for the power spectra and on the specific binning adopted for the flux PDF. We now wish to define a set of parameters describing the overall goodness of the method. For this purpose, we first of all delimit a range of Fourier modes and flux in which it is sensible to compare the statistics obtained via the simulation and the FGPA. In current state-of-the-art high resolution spectra, metal absorption features can increase the 1DPS at  $k \gtrsim 0.1 \text{ s km}^{-1}$  (Lidz et al., 2010, McDonald et al., 2000, 2005, Viel et al., 2013). For this reason, this is typically the maximum  $k$  considered in high-resolution power spectra

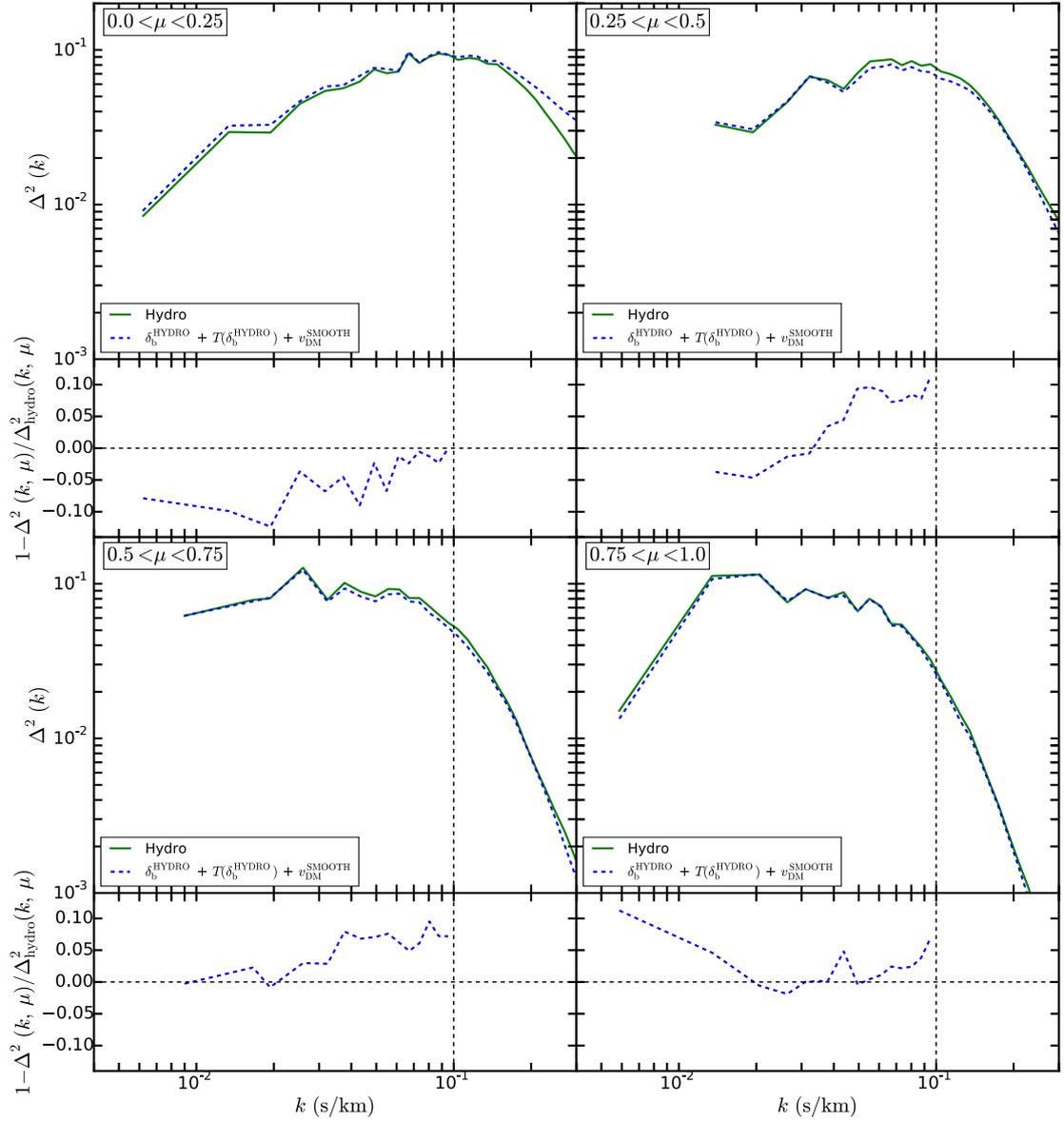


FIGURE 2.4: We show the dimensionless 3DPS  $\Delta^2(k, \mu)$  of the flux given by our reference hydrodynamic simulation (solid green lines) and of the flux computed by imposing a 1-to-1 temperature-density relationship on the baryon density given by the hydrodynamic simulation, and using the Gaussian-smoothed line-of-sight velocities of dark matter instead of baryons (dashed blue lines). We consider 4 bins of  $\mu$ , and show  $\Delta^2(k, \mu)$  as well as the relative difference between the spectra. The dashed vertical line marks the dynamic range considered to compute the accuracy of the FGPA (see text for details). The relative errors plotted show the intrinsic limitations of approximate techniques predicting the Ly $\alpha$  forest through the manipulation of the DM density field given by collisionless simulations.

studies. We shall therefore set our upper limit of the range of  $k$  considered to  $0.1 \text{ s km}^{-1}$ . This upper bound is indicated with the vertical dashed line in the left panels of Figure 2.3. The lower bound of the  $k$ -range considered by us is simply given by the scale of the simulation box. The overall accuracy of the FGPA is assessed by the arithmetic mean of the modulus of the relative error in the range of  $k$  considered:

$$m = \frac{1}{N} \sum_{k < 0.1 \text{ s km}^{-1}} \frac{\left| P_F^{\text{hydro}}(k) - P_F^{\text{FGPA}}(k) \right|}{P_F^{\text{hydro}}(k)} \quad (2.8)$$

where  $N$  is the number of modes in such range. A small value of  $m$  implies a good mean accuracy. Note however that it does not necessarily mean that the accuracy is good *everywhere*. Indeed, a low value of  $m$  can be achieved by a set of points where the relative error is extremely close to zero for many of them but large for just a couple of modes. In other words,  $m$  tells us nothing about the *dispersion* of the relative error around its mean value. To estimate such dispersion, we simply compute the root-mean-square  $s$  of the relative error in the range considered:

$$s^2 = \frac{1}{N} \sum_{k < 0.1 \text{ s km}^{-1}} \left( \frac{\left| P_F^{\text{hydro}}(k) - P_F^{\text{FGPA}}(k) \right|}{P_F^{\text{hydro}}(k)} - m \right)^2 \quad (2.9)$$

The range within which we compute  $m$  and  $s$  is  $0.1 < F < 0.9$ . The upper bound means that we are excluding a range of flux often limited by continuum placement uncertainties (Lee, 2012), whereas the lower bound translates into ignoring flux values susceptible to inaccuracies in modeling optically thick absorbers (Lee et al., 2015). The same analysis is applied to the 3DPS as well, by doing a separate calculation for each bin of  $\mu$  (we consider 4  $\mu$ -bins, evenly spaced between 0 and 1).

The mean accuracy of FGPA at a smoothing length of  $228 \text{ ckpc}$  in reproducing 1DPS and PDF of the flux given by the hydrodynamic simulation is 2%. For the 3DPS, it is between 3% and 5%, depending on the  $\mu$ -bin considered. We stress that these levels of accuracy are obtained employing in the computations the baryon density provided by the hydrodynamic simulation. It means that, regardless how well we create the pseudo density field, this sets our limiting accuracy. To improve it even more, one should come up with more refined ways of reproducing the velocity field and the scatter in the temperature-density relationship.

To sum up, one source of error is considering DM velocities instead of baryonic ones. This is minimized because we looked for the optimal smoothing length for the velocity field. The remaining uncertainty arises from the scatter in the temperature-density relationship which is not captured by the FGPA.

Method	DM Particle Distribution	$(\lambda_G, T_0, \gamma)$	$F_{\text{real}}$ : 3D Power Spectrum and PDF	$F$ : 1D Power Spectrum and PDF
GS+FGPA	✓	✓		
3D-IMS	✓	✓	✓	
1D-IMS	✓	✓	✓	✓

TABLE 2.1: Inputs needed for the different methods considered.

## 2.3 Iteratively Matched Statistics

To better model Ly $\alpha$  forest in collisionless simulations, we developed two novel methods which iteratively match certain Ly $\alpha$  forest flux statistics given as input. The most accurate inputs today come from hydrodynamic simulations, and that is what we use here. We name this technique “Iteratively Matched statistics” (IMS); the two methods are called 3D-IMS and 1D-IMS.

### 2.3.1 3D Iteratively Matched Statistics

The basic idea of 3D-IMS is to compute the flux from a collisionless simulation and match its one- and two-point statistics to a reference hydrodynamic simulation. Because redshift space distortions and thermal broadening make the flux field anisotropic, we for simplicity conduct this matching in real space, where the flux is an isotropic random field. So, in general, one needs a collisionless simulation and a model for the 3D power spectrum and probability distribution function of the flux in real space to apply 3D-IMS. In our case, the model for these statistics is the result of our hydrodynamic simulation. The tabulated 3D power spectrum and PDF of the flux in real space are the inputs of the method, together with the DM particle distribution given by the collisionless simulation and the thermal parameters of the IGM (see Table 2.1). Before going into the details of the procedure, it is worth enumerating the main steps, to better understand the logical flow.

1. As a starting point, the DM density is smoothed with a Gaussian kernel with a smoothing length  $\lambda_G$ . In a situation where the DM was simulated on a coarse grid (e.g. with a PM code),  $\lambda_G$  would be at least as large as the inter-particle simulation. The smoothed field is used to compute the flux in real space within the FGPA, following equations (2.6) and (2.7). We shall call this flux field  $F_{\text{real}}^{\text{DM}}$ .
2. The input real flux dimensionless 3D power spectrum and PDF, taken from the hydrodynamic simulation, are used to calibrate two transformations.

Such transformations are iteratively applied to  $F_{\text{real}}^{\text{DM}}$ , forcing its dimensionless 3D power spectrum and PDF to match the ones given as input. The iterations are implemented until both statistics are matched with high precision.

3. From the resulting pseudo real flux field, a pseudo baryon density field is obtained inverting equations (2.7) and (2.6).
4. The pseudo baryon density is Gaussian-smoothed with a smoothing length equal to the size of a grid cell. As we shall explain later, this step is necessary to remove hot pixels that give rise to non physical density skewers. The smoothed baryon density field is then used to compute flux skewers within the FGPA.

The points just enumerated, which can be visualized as a flow chart in Figure 2.5, give our method its name: Iteratively Matched Statistics (IMS). The prefix 3D stresses that we are matching the dimensionless 3D power spectrum of the flux in real space. Matching this statistics is straightforward, as the  $F_{\text{real}}$  3D power spectrum can be analytically fit by a power law with a Gaussian cutoff (Kulkarni et al., 2015) and is isotropic in redshift space.

On the contrary, reproducing the 3DPS of flux in redshift space would be more complicated, because it is an anisotropic power spectrum. It would require performing transformations in real space after deconvolving redshift space distortions and thermal broadening. Matching the 3D power spectrum of the baryon density would not be optimal either, since it does not exhibit an obvious Jeans cutoff (Kulkarni et al., 2015), being dominated by higher density structures in collapsed halos at small scales. Although these rare dense regions dominate the baryon power spectrum, they contribute negligibly to variations in the Ly $\alpha$  forest flux because the exponentiation of the opacity field maps them to zero. As such, we choose to match the statistics of the real-space flux field, since this is an isotropic field, which is directly related to the observable, that is the flux in redshift space.

We shall now examine the details of each step of the method. We want to remap  $F_{\text{real}}^{\text{DM}}$  to a new field  $F_{\text{real}}^{\text{3D-IMS}}$  with the same dimensionless 3D power spectrum as  $F_{\text{real}}^{\text{HYDRO}}$ . To do this, let us consider the real flux fluctuations  $\delta_{F_{\text{real}}^{\text{DM}}}$  and  $\delta_{F_{\text{real}}^{\text{HYDRO}}}$  in Fourier space. We define  $F_{\text{real}}^{\text{3D-IMS}}$  as  $\delta_{F_{\text{real}}^{\text{3D-IMS}}}(\mathbf{k}) = T(k)\delta_{F_{\text{real}}^{\text{DM}}}(\mathbf{k})$ , where  $T(k)$  is a function tuned to match the dimensionless 3D power spectrum of  $F_{\text{real}}^{\text{HYDRO}}$ . We shall call it ‘‘transfer function’’ and its explicit expression is given by

$$T(k) = \sqrt{\frac{\Delta_{F_{\text{real}}^{\text{HYDRO}}}^2(k)}{\Delta_{F_{\text{real}}^{\text{DM}}}^2(k)}} \quad (2.10)$$

where  $\Delta_X^2(k) = k^3 P_X(k)/2\pi^2$  denotes the dimensionless 3D power spectrum of field  $X$ . Let us point out that in our case it is straightforward to apply equation (2.10), because both  $F_{\text{real}}^{\text{HYDRO}}$  and  $F_{\text{real}}^{\text{DM}}$  sample the same modes, having been built from the same simulation. However, one can apply it also to the more interesting case where  $F_{\text{real}}^{\text{HYDRO}}$  is computed from a small-box hydrodynamic simulation and  $F_{\text{real}}^{\text{DM}}$  from a large-box N-body simulation. This will be discussed into more detail in § 2.5.

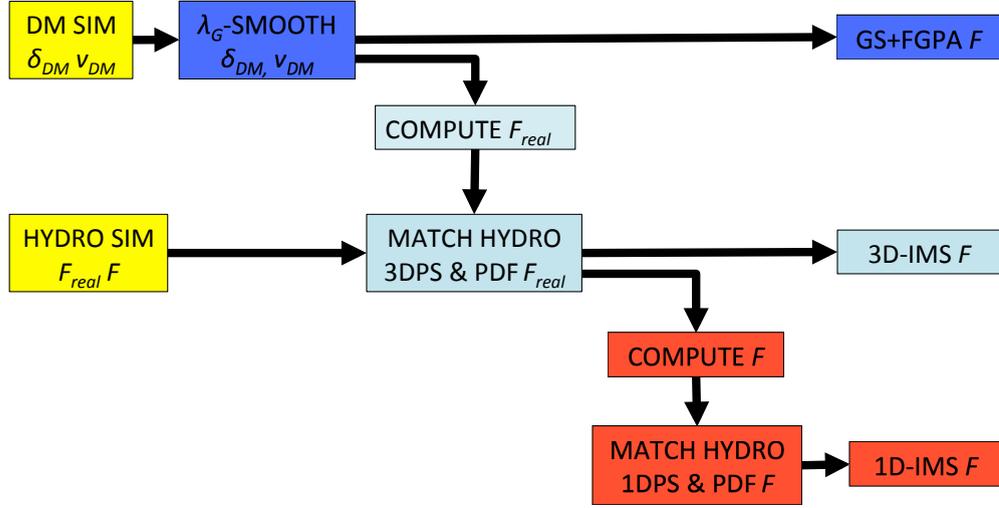


FIGURE 2.5: Flow chart of the methods tested. Yellow boxes are the inputs needed. Blue boxes illustrate the steps of the FGPA applied to the Gaussian-smoothed DM density (GS+FGPA; see § 2.2.1 and § 2.2.2 for details). 3D Iteratively Matched Statistics consists in appending two further steps at the end of GS+FGPA, before computing the flux field. These steps are represented by the cyan boxes. 1D Iteratively Matched Statistics requires to apply two further steps (red boxes) on top of 3D-IMS, just before extracting flux skewers.

At this point, we compute the pseudo real flux field simply as  $F_{\text{real}}^{\text{3D-IMS}}(\mathbf{x}) = \bar{F}_{\text{real}}^{\text{HYDRO}}(1 + \delta_{F_{\text{real}}^{\text{3D-IMS}}}(\mathbf{x}))$ , where  $\bar{F}_{\text{real}}^{\text{HYDRO}}$  is the mean value of the real flux field obtained from the hydrodynamic simulation.

The field  $F_{\text{real}}^{\text{3D-IMS}}$  does not have the same PDF as  $F_{\text{real}}^{\text{HYDRO}}$ . To match the PDF, we use the argument explained by Peirani et al. (2014). We compute the cumulative distribution of both fields and we construct a mapping between the two fluxes by assigning to each value of  $F_{\text{real}}^{\text{3D-IMS}}$  the value of  $F_{\text{real}}^{\text{HYDRO}}$  corresponding to the same percentile in their respective cumulative distributions. We now have a new pseudo real flux field, whose PDF matches by construction the one of  $F_{\text{real}}^{\text{HYDRO}}$ . However, its dimensionless 3D power spectrum is no longer the same as  $\Delta_{F_{\text{real}}^{\text{HYDRO}}}^2(k)$ .

To match both dimensionless 3D power spectrum and PDF of  $F_{\text{real}}^{\text{HYDRO}}$ , we iterate the two transformations. We verified that both 3D power spectrum and PDF converge to their counterparts in the simulation. This is a non trivial result.<sup>5</sup> Convergence occurs between 10 and 20 iterations, after which the improvement in the transfer function at

<sup>5</sup>While seeking the optimal way to match the statistics of the flux fields given by the hydro simulation, we have applied the IMS technique involving also other fields, like  $n_{\text{HI}}$ . Convergence has not occurred in all cases.

every additional iteration is less than 0.3%. It is worth pointing out that every time we match the 3D power spectrum there is no warranty that the new flux field has physically meaningful values, i.e. between 0 and 1. This is indeed the case, so we cannot simply compute  $\delta_b^{3D-IMS}$  from the resulting flux field. This issue is fixed naturally when we match the PDF. Since the distributions are mapped percentile to percentile and  $F_{\text{real}}^{\text{HYDRO}}$  contains obviously only physical values, pixels with negative flux are mapped to small but positive values and pixels with flux larger than one are mapped to values close to but less than 1. It is then fundamental to conclude the iteration process matching the PDF.

At the end of the last iteration, we have the final pseudo real flux field, whose PDF matches by construction the one of  $F_{\text{real}}^{\text{HYDRO}}$ . Since in our model there is a 1-to-1 correspondence between  $\delta_b$ ,  $n_{\text{HI}}$  and  $F_{\text{real}}$ , the PDF of the pseudo baryon density and the hydrogen number density have also converged to an asymptotic distribution. However, in the hydrodynamic simulation there is not such a correspondence, since skewers are not computed within the FGPA. As a result, the PDF of the final  $\delta_b^{3D-IMS}$  does not perfectly match the corresponding field  $\delta_b^{\text{HYDRO}}$  from the reference hydrodynamic simulation. Furthermore, pseudo baryon density skewers present some non physical cuspy overdensities. They arise because, whereas the transformation matching the dimensionless 3D power spectrum of real flux maintains the overall shape of real flux skewers intact, it generally changes the value of the real flux in single pixels. Even a fluctuation as little as  $\sim 10^{-3}$  can increase the real flux at a local minimum above the values of the neighboring pixels, effectively changing their rank-ordering. Since the transformation matching the real flux PDF preserves the rank-ordering of pixels, and because of the exponentiation of equation (2.7), low-flux regions can give rise to large discontinuities in density. These cusps can be eliminated with a Gaussian smoothing. In this way, the density values in neighboring pixels are “blended” together and, as a result, very high values are turned into physical ones. The drawback is that, if we compute the real flux from the smoothed field, it will not have the same PDF as  $F_{\text{real}}^{\text{HYDRO}}$  anymore. A good compromise is adopting the shortest possible length scale for the smoothing, that is the size of one cell of the grid on which we CIC-binned the DM particle distribution. In our case, that corresponds to 28 ckpc. We emphasize here that this last smoothing must always be below the smallest relevant physical scale in the hydrodynamic simulation, which in our context is the Jeans scale, not to considerably affect the resulting statistics.

Running the method for different values of  $\lambda_G$ , we investigate if there is a trend of the accuracy of the various flux statistics. We remind the reader that the initial smoothing serves only as a starting point for the method. In any realistic situation, the value of  $\lambda_G$  is going to be related to the inter-particle separation of the underlying simulation. Indeed, smoothing on a scale smaller than that would make the PDF of the baryon density inaccurate, especially in voids (Rorai et al., 2013), which are the most relevant regions as far as the Ly $\alpha$  forest signal is concerned. The results of our analysis are discussed in § 2.4.

### 2.3.2 1D Iteratively Matched Statistics

The method called 1D-IMS has 3D-IMS as a starting point, on top of which further transformations are applied. Alongside the inputs required by 3D-IMS, one needs to provide a model for the line-of-sight power spectrum and PDF of the flux in redshift space as well (see Table 2.1). Once again, we computed these inputs from the hydrodynamic simulation. After running 3D-IMS, we are left with a real flux field whose dimensionless 3D power spectrum and PDF match the ones of the real flux from the reference hydrodynamic simulation. We then compute the flux in redshift space and apply again the Iteratively Matched Statistics procedure, this time aiming at matching the dimensionless line-of-sight power spectrum and PDF of the flux in redshift space from the hydrodynamic simulation. As in 3D-IMS, we apply two *ad hoc* transformations. Analogously to equation (2.10), we define a transfer function as follows

$$T(k) = \sqrt{\frac{P_{F^{\text{HYDRO}}}^{\text{1D}}(k)}{P_{F^{\text{3D-IMS}}}^{\text{1D}}(k)}}, \quad (2.11)$$

where  $P_{F^{\text{HYDRO}}}^{\text{1D}}(k)$  and  $P_{F^{\text{3D-IMS}}}^{\text{1D}}(k)$  are the line-of-sight power spectra of the flux in redshift space given by the hydrodynamic simulation and obtained after running 3D-IMS respectively. After multiplying the Fourier modes of the fluctuations of  $F^{\text{3D-IMS}}$  by  $T(k)$ , we have a flux field whose dimensionless power spectrum matches  $\Delta_{F^{\text{HYDRO}}}^2(k)$  by construction.

At this point, we match its PDF to the one given by the hydrodynamic simulation exploiting the cumulative distributions, just like in § 2.3.1. We then reiterate the two transformations until we achieve convergence in both 1DPS and PDF. Since these statistics are now matched by construction, it would be interesting to check if the 3D correlations are preserved. We then investigate the trend of the accuracy of the 3DPS as a function of  $\lambda_G$ .

We run 1D-IMS for different values of the initial smoothing length  $\lambda_G$ . The results are discussed in the next section.

## 2.4 Validation of Iteratively Matched Statistics

After implementing the methods described in the previous sections, we assess the accuracy with which we can reproduce the results of the hydrodynamic simulation. In § 3.3.1 we compare the performance of the various techniques in reproducing the skewers of the simulation. In § 2.4.2 and § 2.4.3 we investigate how accurately the statistics of flux are recovered.

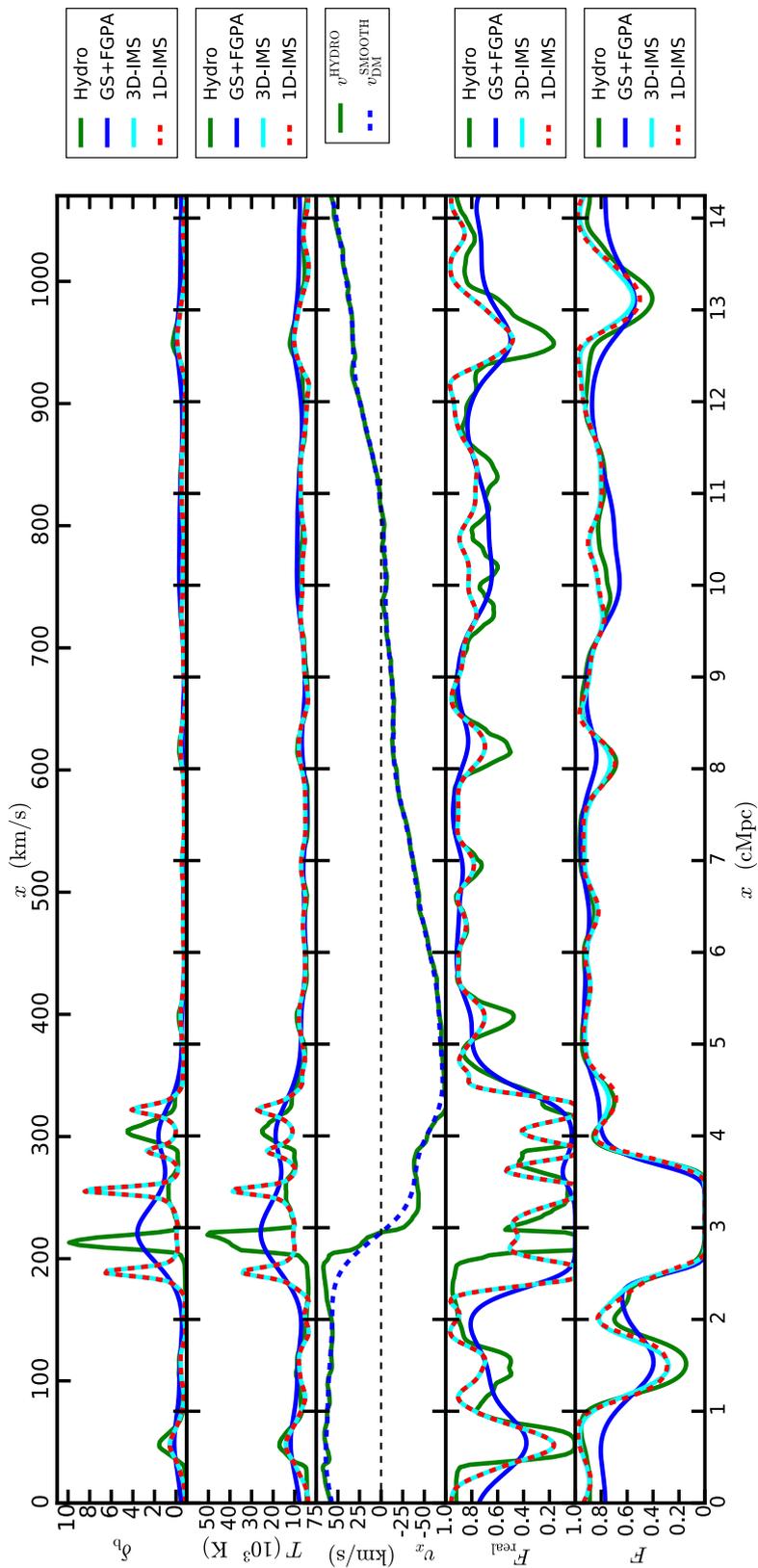


FIGURE 2.6: From top to bottom, baryon density fluctuations, temperature, velocity field, flux in real space and flux in redshift space along a certain skewer. Solid green lines show results from the reference hydrodynamic simulation, solid blue lines refer to GS+FGPA, solid cyan and dashed red lines to 3D-IMS and 1D-IMS respectively. 1D-IMS consists in matching the dimensionless line-of-sight and the PDF of the flux in redshift space on top of the results given by 3D-IMS, so these two methods differ only for this quantity. The dashed blue line in the third panel from top represents the Gaussian-smoothed line-of-sight velocities of dark matter. For all methods, we plotted the curves corresponding to the optimal value of the initial smoothing scale. The skewers obtained in all cases are consistent with one another.

### 2.4.1 Skewers

In Figure 2.6 we show different quantities along one skewer as an example. From top to bottom, we plot the baryon density fluctuations, temperature, velocity field, flux in real space and flux in redshift space. In all panels, solid green lines refer to the skewers extracted from the hydrodynamic simulation, solid blue lines to the ones obtained through GS+FGPA, and solid cyan and dashed red lines to 3D-IMS and 1D-IMS, respectively. Each curve corresponds to the optimal smoothing length for the respective method. The dashed blue line in the third panel from the top refers to the line-of-sight velocities of DM, Gaussian-smoothed at  $\lambda_G = 228$  ckpc. This is the velocity field used in all approximate methods to compute all quantities above (see appendix A).

The 1D-IMS technique has 3D-IMS as its starting point, and differs from it for two additional transformations to match the dimensionless 1DPS and the PDF of the flux in redshift space with the results from the hydrodynamic simulation. Therefore, the flux in real space, and consequently baryon density fluctuations and temperature fields, are the same as in 3D-IMS.

We can see that all methods result in skewers that trace those of the hydrodynamic simulation very well, and are also consistent with one another. This means that not only is IMS able to reproduce the statistics of the Ly $\alpha$  forest correctly, but it also generates reasonable mock skewers. This did not obviously have to be the case. For example, the method LyMAS (Peirani et al., 2014) is designed to match the 1DPS and PDF of the flux from hydrodynamic simulations as well, but only the more complex version of LyMAS, which involves two additional transformations, produces reasonable-looking skewers. In our techniques, the mappings guarantee that both statistics and flux skewers are reproduced accurately.

Furthermore, not only is the flux accurately reproduced, but also the other quantities plotted in Figure 2.6. The biggest difference between IMS methods and GS+FGPA is that the former better reproduces high and narrow density peaks, like the ones around 3 cMpc and 4 cMpc in Figure 2.6.

Conversely, this is not always the case for smaller overdensities, such as the one around 0.7 cMpc in Figure 2.6, where IMS produces a lower density peak than in the hydro. Note however that at this location the flux in real space is still much more accurate with our methods, since they are designed to match its 3D power spectrum. Small differences in  $F_{\text{real}}$  can easily yield large differences in density because of the exponential in equation (2.7). Such differences persist also in temperature, which in our context is connected to the density through a pure power law. Flux skewers in redshift space appear to be more similar among the various methods, since the convolution of real flux with the velocity field and thermal broadening tends to smooth out the differences.

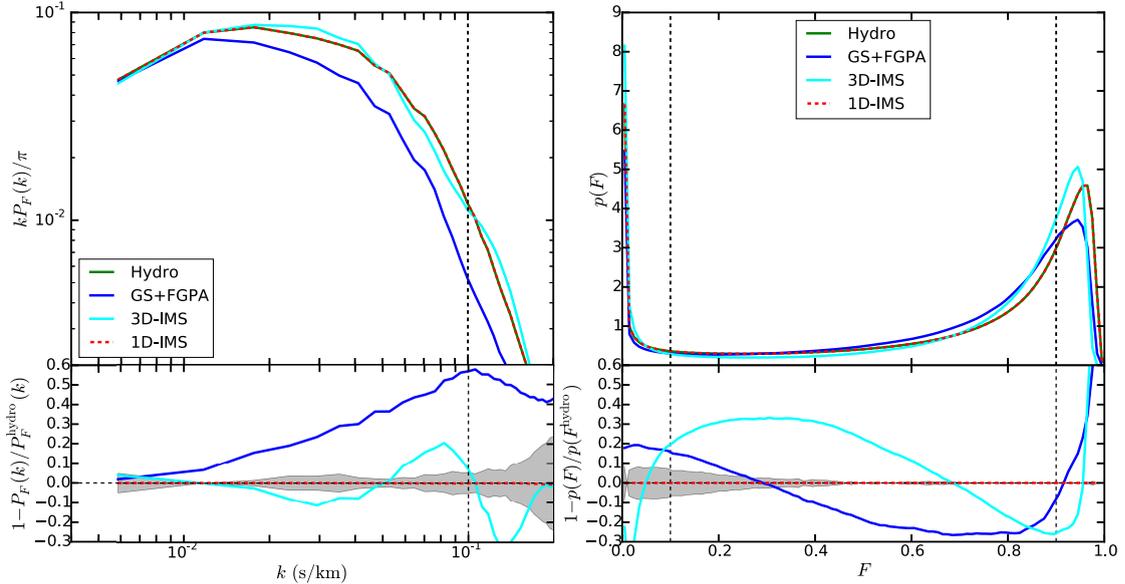


FIGURE 2.7: *Top panels:* Line-of-sight power spectrum (left) and PDF (right) of flux, given by the reference hydrodynamic simulation (solid green line), GS+FGPA (solid blue line), 3D-IMS (solid cyan line) and 1D-IMS (dashed red line). The results plotted refer to runs with initial smoothing length 228 ckpc. *Bottom panels:* On the left, relative difference between the 1D power spectrum obtained through the different methods tested and the one given by the hydrodynamic simulation. On the right, analogous plot for the PDF. In both panels, the shaded area represents the region within which the relative difference is smaller than the one obtained applying a 1-to-1 temperature density relationship to the baryon density given by the hydrodynamic simulation and using the Gaussian-smoothed line-of-sight velocities of dark matter instead of baryons. In all panels, the dashed vertical lines delimit the dynamic range considered to compute the accuracy. Horizontal dashed black lines mark the zero difference level and are meant to guide the eye. Our methods reproduce the line-of-sight better than GS+FGPA. In particular, 1D-IMS matches both power spectrum and PDF by construction.

## 2.4.2 Comparison of the Methods

All methods we considered have GS+FGPA as their starting point, with  $\lambda_G$  as a free parameter. We now compare the flux statistics given by each method with the ones from the hydrodynamic simulation, varying  $\lambda_G$  in the range 0 – 570 ckpc, in steps of 57 ckpc. The dependence on this parameter of the accuracy in reproducing the various statistics is different for each method. It is generally possible to identify an optimal value of  $\lambda_G$  for a given method at matching a certain statistic, but this may not be optimal for all flux statistics.

In the top panels of Figure 2.7 we compare 1DPS and PDF of the flux in redshift space given by the hydrodynamic simulation (solid green line) to the approximate methods considered. Solid blue, solid cyan and dashed red lines refer to GS+FGPA, 3D-IMS and 1D-IMS, respectively. We plotted the curves corresponding to  $\lambda_G = 228$  ckpc for each method. Since 1D-IMS is designed to match dimensionless 1DPS and PDF of the flux, solid green and dashed red lines are indistinguishable. We can see that 3D-IMS reproduces well both 1DPS and PDF, whereas GS+FGPA does not recover well the

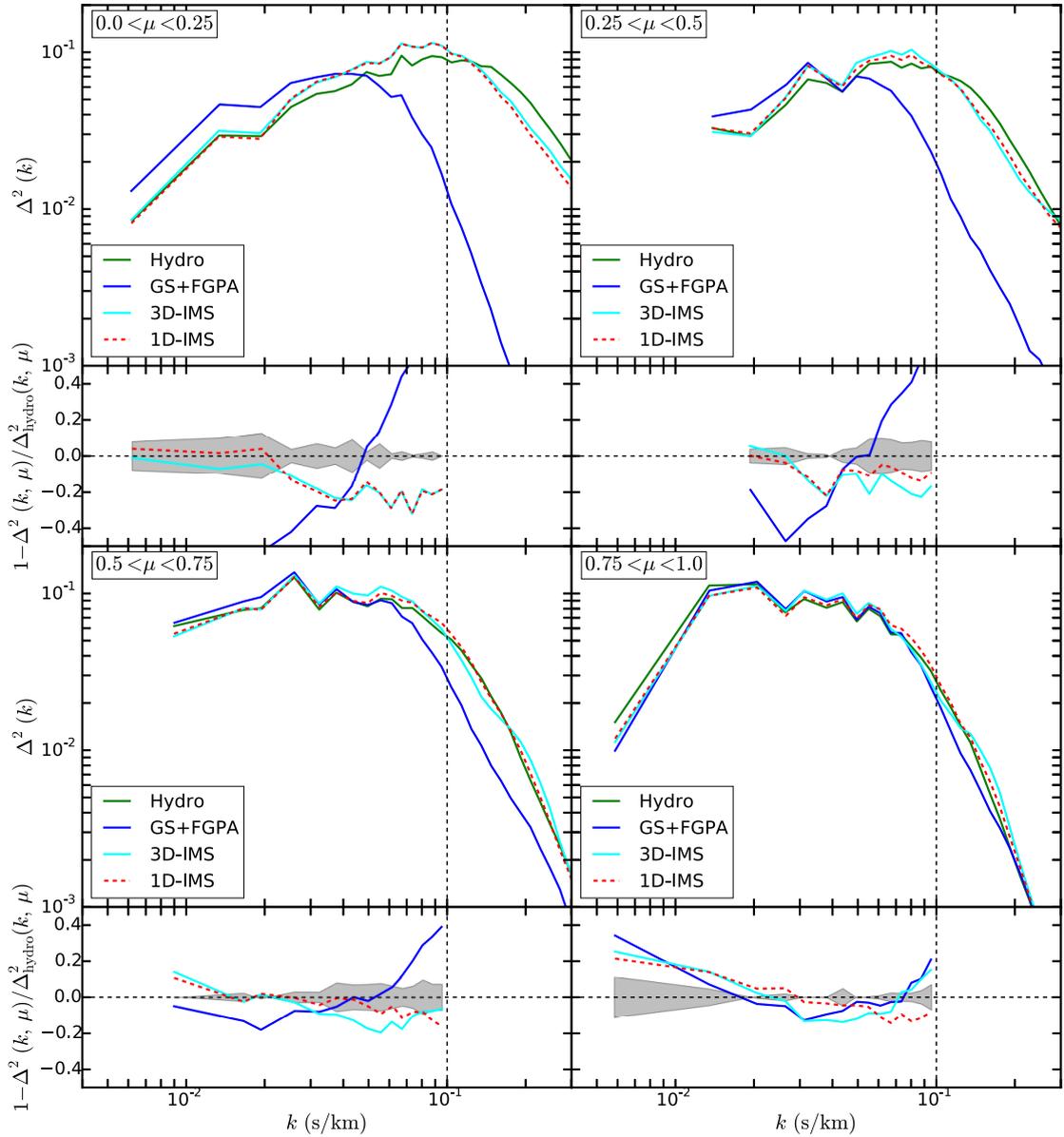


FIGURE 2.8: We show the dimensionless 3D power spectrum  $\Delta^2(k, \mu)$  of the flux given by our reference hydrodynamic simulation (solid green lines), by GS+FGPA (solid blue lines), 3D-IMS (solid cyan lines) and 1D-IMS (dashed red lines). The results plotted refer to runs with initial smoothing length  $\lambda_G = 228$  ckpc. We considered 4 bins of  $\mu$ . For each one of them, there are two panels. The upper one shows  $\Delta^2(k, \mu)$  versus  $k$  in the  $\mu$ -bin considered, the lower one the relative difference between the spectra. In all panels, the dashed vertical lines delimit the dynamic range considered to compute the accuracy. Horizontal dashed black lines mark the zero difference level and are meant to guide the eye. Shaded areas represent the regions within which the relative difference is smaller than the one obtained applying a 1-to-1 temperature density relationship to the baryon density given by the hydrodynamic simulation and using the Gaussian-smoothed line-of-sight velocities of dark matter instead of baryons. Our methods perform better than GS+FGPA in all  $\mu$ -bins.

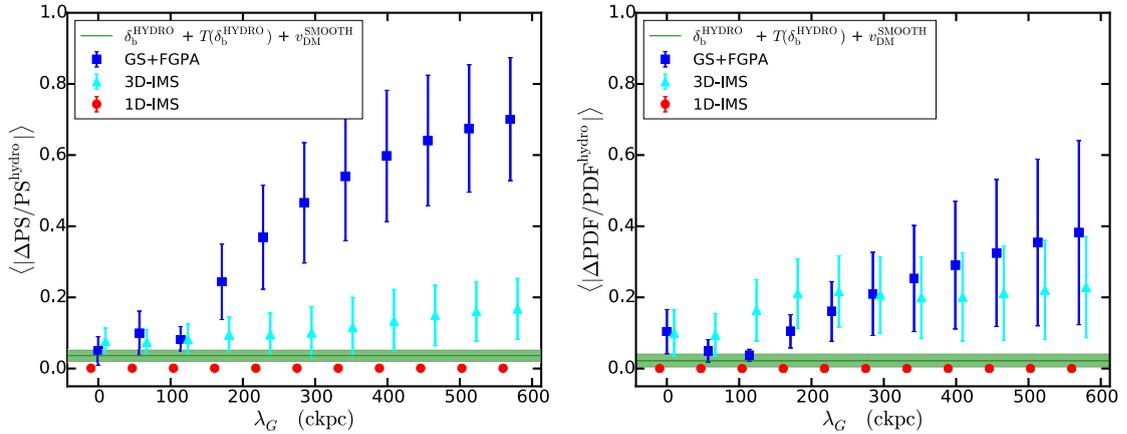


FIGURE 2.9: Accuracy of the different methods tested in reproducing the flux dimensionless line-of-sight power spectrum (left panel) and PDF (right panel) given by the reference hydrodynamic simulation, as a function of the initial smoothing length  $\lambda_G$ . Markers indicate the mean values of the accuracy, while error bars represent the root-mean-square of the accuracy in the dynamic ranges considered. Blue squares refer to GS+FGPA, cyan triangles to the 3D-IMS and red circles to the 1D-IMS. An offset of  $\pm 10$  ckpc has been applied to 3D-IMS and 1D-IMS markers to make the plot more readable. The horizontal green line shows the mean accuracy obtained by applying a 1-to-1 temperature-density relationship to the baryon density field and using the Gaussian-smoothed line-of-sight velocities of dark matter baryons. The green band represents the root-mean-square of the accuracy in this case. Our methods are overall more accurate and less dependent on the initial smoothing scale than GS+FGPA.

1DPS. The lower panels make the comparison quantitative, showing the relative error of each method at recovering the results of the reference simulation. The gray shaded area represents the region within which the relative difference is smaller than the one obtained applying the FGPA to the baryon density given by the hydrodynamic simulation and using the smoothed DM velocity field, as explained in section 2.2.3. As previously said, this sets the limits on the accuracy due to adopting the DM-smoothed velocity field and neglecting the scatter in the temperature-density relationship of the IGM.

Figure 2.7 then tells us that 1D-IMS is able to recover the information lost with these approximations by construction, since it was forced to match the redshift space 1DPS and PDF of the hydrodynamic simulation. In contrast, the flux PDF given by 3D-IMS does not appear very accurate, perhaps even erroneously suggesting a flaw in the method. This is not the case, as 3D-IMS matches the PDF of the flux in real space, whereas in the right panel we are considering the PDF of the flux in redshift space. Although the relative error of the 3D-IMS PDF is as large as 30% at  $F = 0.2$ , the average accuracy is 15%. When the optimal value of  $\lambda_G$  is used for 3D-IMS (57 ckpc), the PDF is reproduced with an average accuracy of 8%. The variability of the accuracy at different flux values is not too surprising, because in the last step of 3D-IMS we smooth the pseudo baryon density field to remove hot pixels and this impacts the accuracy of the corresponding flux PDF.

Figure 2.8 shows the 3DPS given by the simulation and the various methods at  $\lambda_G =$

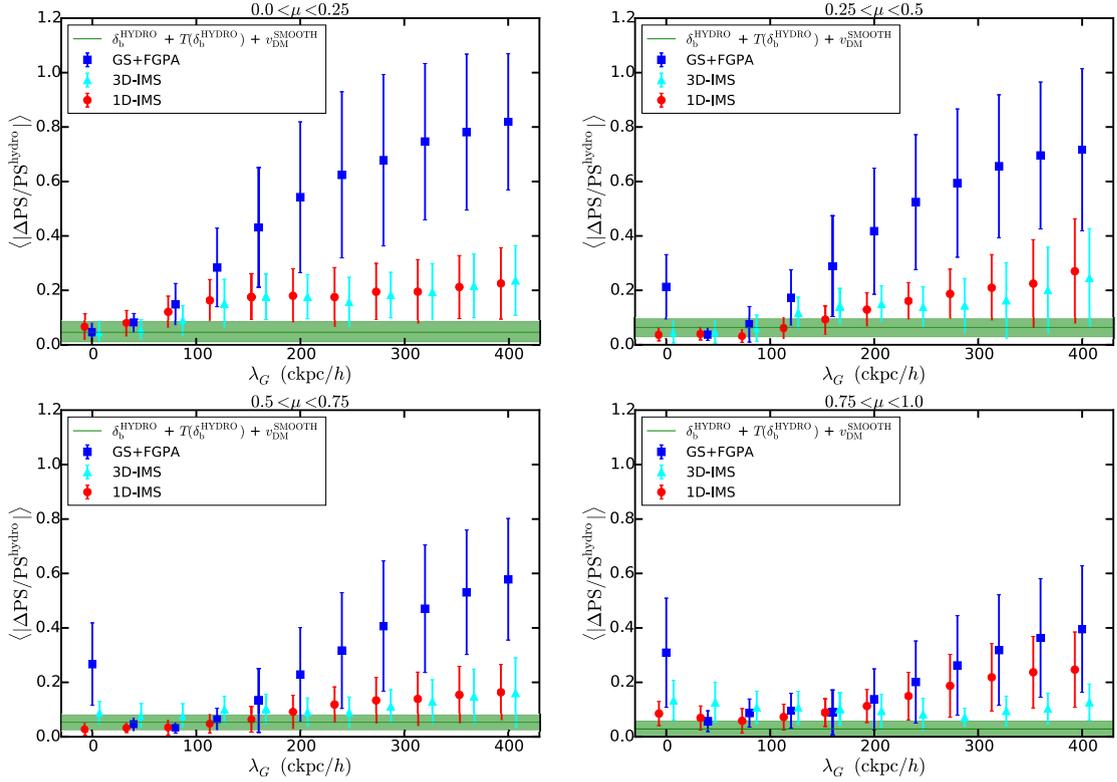


FIGURE 2.10: Accuracy of the different methods tested in reproducing the dimensionless 3D power spectrum  $\Delta^2(k, \mu)$  of the flux given by the reference hydrodynamic simulation, as a function of the initial smoothing length  $\lambda_G$ . Each panel shows the results obtained for a different bin of  $\mu$ . In all panels, markers indicate the mean values of the accuracy, while error bars represent the root-mean-square of the accuracy in the dynamic ranges considered. Blue squares refer to GS+FGPA, cyan triangles to the 3D-IMS and red circles to the 1D-IMS. An offset of  $\pm 10$  ckpc has been applied to 3D-IMS and 1D-IMS markers to make the plot more readable. The horizontal green lines show the mean accuracy obtained by applying a 1-to-1 temperature-density relationship to the baryon density field and using the Gaussian-smoothed line-of-sight velocities of dark matter instead of baryons. The green bands represent the root-mean-square of the accuracy in this case. In all  $\mu$ -bins, our methods are overall more accurate and less dependent on the initial smoothing scale than GS+FGPA.

228 ckpc, as well as the relative error in matching this statistic. The color coding is the same as in Figure 2.7. Each panel refers to a different  $\mu$ -bin of the 3DPS. In all bins, the accuracy of 3D-IMS and 1D-IMS looks on average comparable to the limit set by the FGPA.

In Figure 2.8 one can clearly see that GS+FGPA in the top-left panel ( $0.0 < \mu < 0.25$ , farthest from the line-of-sight) does not match the hydrodynamic result as well as in the other panels. This is due to the different effects at work at different directions from the line-of-sight. For very transverse modes ( $0.0 < \mu < 0.25$ ) the behavior of baryons is mostly influenced by the filtering scale. Whereas for modes that are parallel to the line-of-sight ( $0.75 < \mu < 1.0$ ) the effect of the Jeans filtering is degenerate with thermal broadening and redshift space distortions (Rorai et al., 2013). As a result, in the bin closest to the line-of-sight ( $0.75 < \mu < 1.0$ ) one can compensate a bad choice of  $\lambda_G$  with

an accurate description of the thermal state of the IGM, whose parameters ( $T_0$  and  $\gamma$ ) are obtained by fitting outputs of the hydrodynamic simulation. But it is not possible to apply this correction in the bin with modes most transverse to the line-of-sight, where the effect of the filtering scale dominates the shape of the 3D power. Furthermore, 3D-IMS is superior to GS+FGPA to the extent that it also matches the dimensionless 3D power spectrum of flux in real space by construction.

Regarding 1D-IMS, it might seem puzzling that it does not perfectly match the result of the hydrodynamic simulation in the bin closest to the line-of-sight, since 1D-IMS is forced to reproduce the dimensionless line-of-sight power spectrum by construction. However, the 3DPS in the bin closest to the line-of-sight is not exactly the 1DPS. Indeed, the power spectrum in that bin considers all flux fluctuations whose wavevector forms an angle with the line-of-sight such that its cosine is between 0.75 and 1. This is a 3D region in redshift space. In the case of the 1DPS, the situation is much different, since one considers flux fluctuations exclusively along the direction of the line-of-sight. Therefore, the 1DPS and 3D power for the bin closest to the line-of-sight are not exactly the same, and matching the 1DPS by construction does not guarantee a perfect match in this bin of the 3DPS. It is true, however, that the agreement should be much better if one considers  $\mu$  values progressively closer to being parallel to the line-of-sight ( $\mu = 1$ ), which we have verified directly.

### 2.4.3 Accuracy versus Smoothing Length

The best simulations reproducing BOSS<sup>6</sup>/DESI<sup>7</sup>-like surveys have a mean inter-particle separation of  $\sim 400$  ckpc. As we have already mentioned, Gaussian-smoothing below the the inter-particle separation has a negligible effect. Therefore, it is of great interest to test the accuracy of GS+FGPA and our methods at different values of the smoothing length, including  $\lambda_G > 300$  ckpc. For this purpose, we compute the mean and root-mean-square of the accuracy for each value of  $\lambda_G$ , as explained in section 2.2.3.

We show the results of our analysis for the 1DPS and PDF in Figure 2.9, in the left and right panels, respectively. The information given by Figure 2.7 is here condensed in three points, one for each method, at the corresponding value of  $\lambda_G$ . Blue squares represent the mean accuracy  $m$  of GS+FGPA and the corresponding error bars the root-mean-square  $s$ , as defined in equations (2.8) and (2.9), respectively. Likewise, cyan triangles and red circles refer to 3D-IMS and 1D-IMS, respectively. The information encoded by the gray shaded areas in Figure 2.7, which shows the limitations of the FGPA, is represented by the green band in Figure 2.9. Hence, the green line shows the mean accuracy given by the FGPA implemented as in section 2.2.3 and the shaded green area delimits  $1\text{-}\sigma$  deviations from this mean. There is no dependence on  $\lambda_G$  in

<sup>6</sup>Baryon Oscillation Spectroscopic Survey (Dawson et al., 2013)

<sup>7</sup>Dark Energy Spectroscopic Instrument (DESI Collaboration et al., 2016)

this case, because the flux within the FGPA is computed from the baryon density field given by the hydrodynamic simulation and not from a Gaussian-smoothed DM density field. Figure 2.10 shows the results for the 3DPS, with the same format as Figure 2.9. Each panel of Figure 2.10 refers to a different  $\mu$ -bin.

As expected, Figures 2.9 and 2.10 show that GS+FGPA is strongly dependent on  $\lambda_G$ . The optimal value appears to be around 114 ckpc for the 1DPS and between 57 ckpc and 114 ckpc for the PDF. Around these values,  $m \approx 7\%$  for the 1DPS and  $m \approx 4\%$  for the PDF, as can be seen from the blue points in Figure 2.9.<sup>8</sup>

The trend of the accuracy of GS+FGPA in reproducing the 3DPS is similar to the one of the 1DPS (blue squares in Figure 2.10). The mean accuracy achieved in the different  $\mu$  bins at the optimal scale for the 3DPS ( $\lambda_G = 57$  ckpc) is around 4%. Remarkably, the accuracy of GS+FGPA for all statistics approaches the limit set by the FGPA, as long as the “correct” smoothing length is chosen. Since the optimal scales for the statistics considered vary up to a factor of two, one should decide in advance whether to prioritize 1DPS, 3DPS or PDF. For  $\lambda_G \gtrsim 171$  ckpc, the accuracy of all statistics gets worse than  $\sim 20\%$ . Moreover, the error bars are very large for smoothing scales  $\gtrsim 200$  ckpc. As such, it can be much worse than the mean in certain ranges of  $k$ -modes and flux. We also note that the performance of GS+FGPA degenerates as one moves farther from the line-of-sight, as previously discussed in the context of Figure 2.8.

Even for initial smoothing lengths  $\gtrsim 200$  ckpc, 3D-IMS yields  $m < 20\%$  for 1DPS and PDF, as shown by the cyan triangles in Figure 2.9, performing significantly better than the Gaussian method for these large smoothing lengths. At smaller smoothing lengths, 3D-IMS is basically as accurate as GS+FGPA. The accuracy in the 1DPS is better than in the PDF. This is not so surprising since, as we already pointed out, we ended our iterations matching the PDF of the flux in real space, and not redshift space, of the hydrodynamic simulation. Figure 2.10 shows that 3D-IMS does a remarkable job of reproducing the 3DPS, with an accuracy comparable to the FGPA at small  $\lambda_G$ , and still around 7% even for initial smoothing lengths as large as 500 ckpc. Moreover, the accuracy of 3D-IMS is only weakly dependent on  $\lambda_G$ , and it performs much better than GS+FGPA for large smoothing lengths.

By construction 1D-IMS matches the 1DPS and PDF resulting in  $m \approx 0.03\%$  independent of  $\lambda_G$  (red circles in Figure 2.9). Figure 2.10 shows that 1D-IMS preserves 3D correlations, yielding  $m = 3.3\%$  for the 3DPS in the best case (57 ckpc in the bin  $0.25 < \mu < 0.5$ ) and  $m = 27\%$  in the worst one (570 ckpc in the bin  $0.25 < \mu < 0.5$ ). In

<sup>8</sup>The minimum in the accuracy of the 1DPS at  $\lambda_G = 0$  ckpc would suggest that the best result is obtained without smoothing the DM density at all. If we apply no smoothing, we are actually limited by the resolution of the simulation. In our context, the DM was solved using a PM code on a grid with size of 28 ckpc and that also corresponds to the inter-particle separation. The DM density was also implicitly smoothed by the CIC kernel on that scale, which is thus the effective smoothing length corresponding to  $\lambda_G = 0$ . There is hence nothing peculiar about the point at  $\lambda_G = 0$ . Furthermore, the overall accuracy corresponding to this value is actually similar to the value obtained at  $\lambda_G = 57$  ckpc.

all bins, 1D-IMS is as accurate as Gaussian smoothing at small smoothing lengths and it performs better than this method for  $\lambda_G \gtrsim 142$  ckpc.

The accuracy of 1D-IMS improves as the  $\mu$ -bin considered approaches the line-of-sight. It performs worse than 3D-IMS in the bin closest to it ( $0.75 < \mu < 1.0$ ). This is counter-intuitive, but we point out that the most parallel bin takes into account correlations in a 3D region of space and is thus conceptually distinct from the 1DPS, which 1D-IMS matches by construction. When recovering the 3DPS in the bin closest to the line-of-sight ( $0.75 < \mu < 1.0$ ), it is still more important to correctly reproduce the 3D correlations rather than the correlations along the line-of-sight. This is why 3D-IMS looks better than 1D-IMS close to the line-of-sight. Similar to 3D-IMS, the accuracy of 1D-IMS is only weakly dependent on the smoothing length, and much better than GS+FGPA for large smoothing lengths.

Among the methods considered in this work, 1D-IMS seems to perform the best. Indeed, it perfectly matches the 1DPS and PDF (by construction) and reproduces the 3DPS with a good accuracy. If one is primarily interested in the 3DPS, 3D-IMS may be more suitable, since it yields the best accuracy in this statistic, although the differences with 1D-IMS are small. The drawback of 3D-IMS is the relative inaccuracy in the 1DPS and PDF compared to 1D-IMS, which matches these statistics by construction. The Gaussian smoothing can recover all statistics as well as 3D-IMS, provided the appropriate  $\lambda_G$  is adopted. In particular, the errors in estimating the 3DPS in the bin farthest from the line-of-sight ( $0.0 < \mu < 0.25$ , top-left panel in Figure 2.10) are larger than  $\sim 20\%$  for  $\lambda_G \gtrsim 171$  ckpc. For comparison, 3D-IMS and 1D-IMS achieve  $\lesssim 10\%$  accuracy for  $\lambda_G \lesssim 228$  ckpc in the aforementioned  $\mu$ -bin. This means that our methods are able to recover information that gets otherwise lost when performing a Gaussian smoothing. They are accurate and computationally cheap ways to reproduce the statistics of the Ly $\alpha$  forest, which have promise for future modeling and data analysis.

We have applied the same analysis described so far also to two snapshots at redshifts  $z = 2$  and  $z = 4$ , respectively. The accuracy of all methods are comparable with the results obtained at  $z = 3$ , meaning that the techniques tested are robust in the range  $2 < z < 4$ . We have also verified that, with  $256^3$  resolution elements, the accuracy of all methods differs by  $\sim 1\%$  from the values obtained with our reference simulations, in most of the range of  $\lambda_G$ . This means that the accuracy of the methods has converged in our study.

When applying our methods, the choice of the initial smoothing length for the DM density is set by the inter-particle separation of the simulation adopted. If this is smaller than the optimal smoothing length, then one should smooth the DM density at the optimal  $\lambda_G$ . Otherwise, the best one can do is adopting a smoothing length of the order of the inter-particle separation. In any case, the smoothing scale for the DM line-of-sight velocities can be larger than the value adopted for the DM density. Indeed, the velocity field itself is smooth in voids (Aragon-Calvo & Szalay, 2013, van de Weygaert

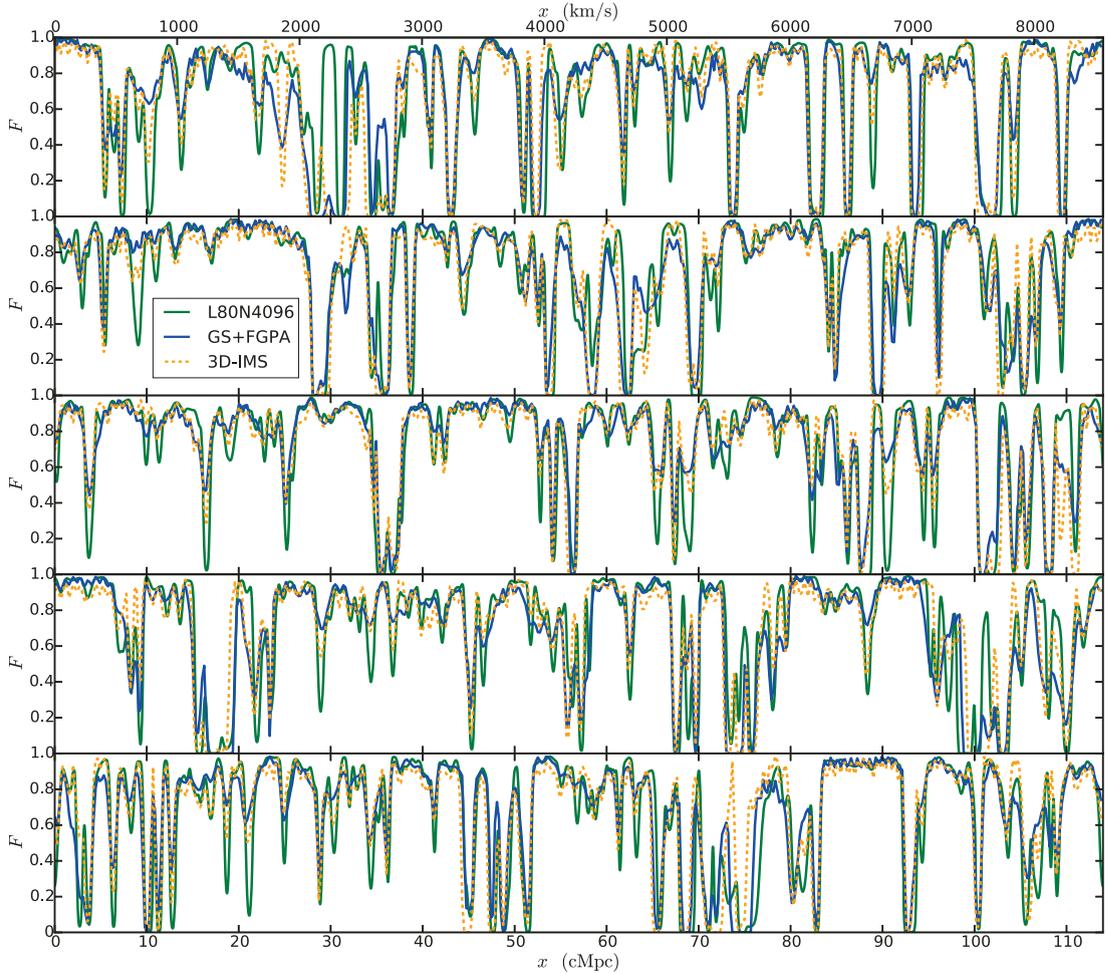


FIGURE 2.11: Sample of five flux skewers obtained through different methods: L80N4096 hydrodynamic simulation with box size 114 cMpc and  $4096^3$  resolution elements (solid green line), which we assume to be the “truth”, GS+FGPA with smoothing length 228 cMpc (solid blue line) and 3D-IMS (dashed orange line). The skewers obtained through all methods are consistent with one another.

& van Kampen, 1993) and these are the most relevant regions for our study, as the exponentiation in equation (2.7) suppresses large overdensities. In our analysis, we kept the smoothing length for the DM velocity field fixed to 228 cMpc, which is the value that yields the best overall accuracy in reproducing the flux statistics considered (see appendix A). In this way, we focused on the impact of the initial smoothing of the DM density field on the accuracy of the methods. Due to our choice of optimizing the smoothing length of the DM velocity field, the errors quoted for the different methods are minimized. However, even if we did not use the optimal  $\lambda_G$  for the velocity, the trend of the accuracy versus the smoothing length of the DM density field would be unaffected, as well as the the rank ordering of the accuracy of the various techniques investigated (see appendix A for a detailed discussion).

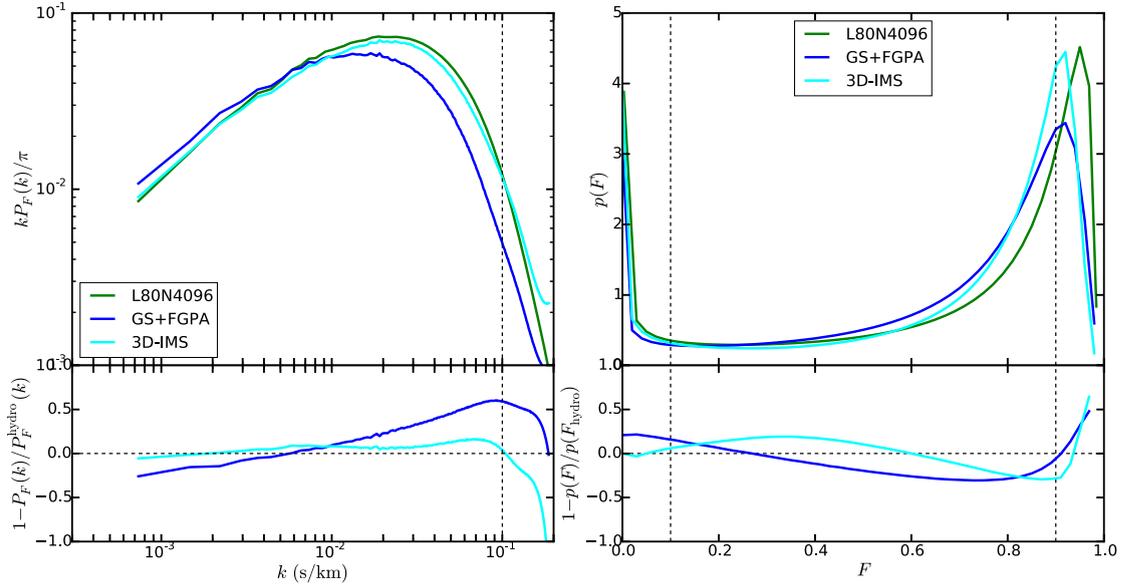


FIGURE 2.12: Flux line-of-sight power spectrum (left) and PDF (right) given by L80N4096 hydrodynamic simulation (solid green line), assumed to be the “truth”, and obtained applying GS+FGPA (solid blue line) and 3D-IMS (solid cyan line) to the DM-only simulation. In all panels, the dashed vertical lines delimit the dynamic range considered to compute the accuracy. All results plotted refer to runs with initial smoothing length 228 ckpc. This is the smallest smoothing allowed by the resolution of the simulation, so the Gaussian smoothing is already optimized. Nevertheless, it is very inaccurate in recovering the line-of-sight power spectrum, meaning that 3D-IMS is certainly superior.

## 2.5 Large-Volume Collisionless Simulation

We want to check if our methods still perform well when applied to an actual N-body run, with a larger box than our calibrating hydrodynamic simulation. In fact, in the previous sections we have validated our methods extracting all relevant fields from the same hydrodynamic simulation. However, the purpose of approximate methods is avoiding expensive hydrodynamic simulations. In practice, one would assume a certain model for the flux statistics and apply our techniques to a large-box low-resolution DM-only run. In this way, one would be able to probe large scales and at the same time accurately describe the small-scale physics thanks to our iterative procedure.

We consider the snapshot at redshift  $z = 3$  of a Gadget DM-only run with a box size of 114 cMpc and  $512^3$  particles. We CIC-bin the particle positions and velocities on a grid with  $512^3$  elements, to get the density and velocity fields. To mock baryonic pressure, we smooth both fields with a length scale of  $\lambda_G = 228$  ckpc, very close to the cell size (223 ckpc). This is the smallest smoothing length one can choose to have a non-negligible effect on the density and velocity fields.<sup>9</sup> We then apply GS+FGPA in § 2.2.1.

<sup>9</sup>The smoothing scale adopted here is also very close to the optimal value for the velocity field determined when validating our method with the smaller hydrodynamic simulation (see appendix A).

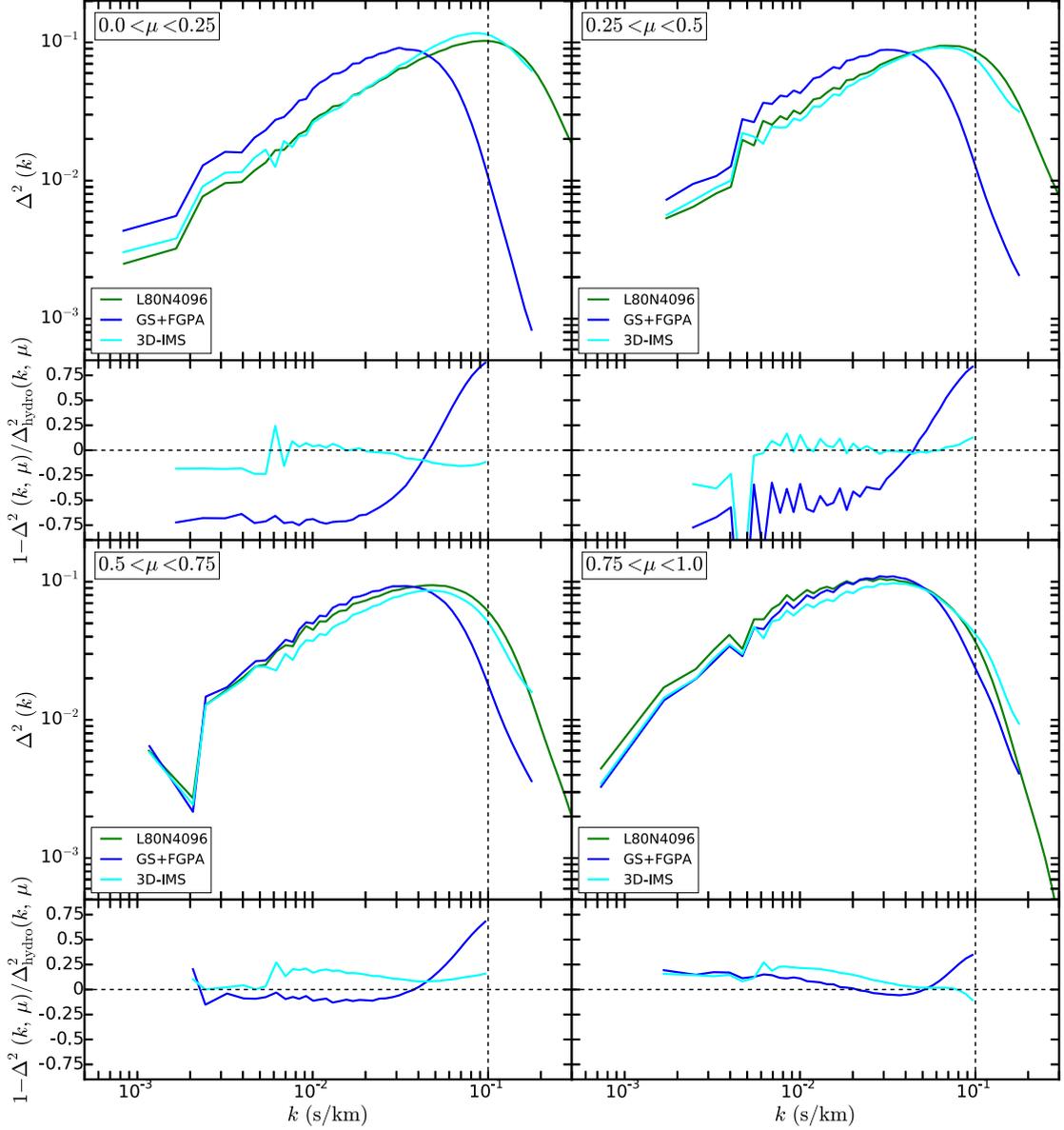


FIGURE 2.13: Dimensionless 3D power spectrum  $\Delta^2(k, \mu)$  of the flux given by L80N4096 hydrodynamic simulation (solid green lines), assumed to be the “truth”, and obtained applying GS+FGPA (solid blue lines) and 3D-IMS (solid cyan lines) to the DM-only simulation. The results plotted refer to runs with initial smoothing length  $\lambda_G = 228$  ckcpc. We considered 4 bins of  $\mu$ . Each panel shows  $\Delta^2(k, \mu)$  versus  $k$  in the  $\mu$ -bin considered. In all panels, the dashed vertical lines delimit the dynamic range considered to compute the accuracy. Whereas 3D-IMS is accurate and its performance does not depend strongly on the  $\mu$ -bin considered, GS+FGPA fails at reproducing the true  $\Delta^2(k, \mu)$  in the  $\mu$ -bins closer to the line-of-sight ( $0.0 < \mu < 0.25$  and  $0.25 < \mu < 0.5$ ).

To apply 3D-IMS, we need an input model for the 3D power spectrum and PDF of the flux in real space. These statistics are computed from the flux in real space given by our hydrodynamic simulation. However, its box size is smaller than the one of the N-body simulation. This does not affect the computation of the PDF, but it poses some problems with the 3D power spectrum, as the hydrodynamic simulation lacks the large modes which are present in the DM-only simulation. To generalize the 3D-IMS method, we construct the transfer function defined in equation (2.10) as follows. First of all, we fit the 3D power spectrum of the flux in real space given by the hydrodynamic simulation with the formula provided by Kulkarni et al. (2015). Then, at every iteration of 3D-IMS, we define the transfer function applying equation (2.10) for  $k$ -modes larger than the fundamental mode  $k_f^{\text{HYDRO}}$  of the hydrodynamic simulation. For the modes smaller than  $k_f^{\text{HYDRO}}$  we set the transfer function to a constant, equal to the value assumed at  $k_f^{\text{HYDRO}}$ . In this way, we have a continuous transfer function, which simply rescales by a constant the large-scale modes probed only by the DM-only simulation. This guarantees that any peculiar large-scale feature in the DM-only simulation (e.g. BAO signal), will not be affected.

Modeling the 1DPS of the flux in redshift space presents similar issues. Once again, one should come up with a method to estimate the large-scale Fourier modes without actually running a calibrating large-box hydrodynamic simulation. It would then be desirable to adopt an approach analogous to the one described for the 3D power spectrum of the real flux field. Unfortunately, there is no simple fitting function available for the 1DPS of the flux in redshift space which would grant the high level of accuracy we are aiming for. Therefore, we are not applying 1D-IMS to the DM-only simulation. Nevertheless, following Kulkarni et al. (2015), future work may provide a fitting function for the flux 1DPS as well.

To assess the accuracy of the various techniques in this test, we shall compare the results of each method with the flux statistics obtained from a Nyx hydrodynamic run with a box size of 114cMpc and  $4096^3$  resolution elements, which we assume to be the “truth”. This simulation, to which we shall refer as “L80N4096”, covers the largest modes present in the DM-only simulation and, at the same time, has the same resolution limit ( $\sim 28$  ckpc) as the small calibrating hydrodynamic simulation, thus resolving the Jeans scale.

In Figure 2.11 we show a sample of five skewers extracted from the hydrodynamic simulation (solid green line) and obtained through GS+FGPA and 3D-IMS (solid blue and dashed orange lines, respectively). By visual inspection, all skewers look consistent with one another.

In the upper-left panel of Figure 2.12 we show the 1DPS of the flux given by the hydrodynamic simulation (solid green line) and obtained applying GS+FGPA and 3D-IMS to the DM-only run (solid blue and cyan lines, respectively). In the lower-left panel, we plot the relative difference of the flux 1DPS obtained in each case, with respect to the

results of L80N4096. In the right panels, we show analogous plots for the flux PDF, following the same color coding. Applying the same analysis outlined in previous sections, we find out that GS+FGPA (3D-IMS) recovers the flux 1DPS with  $m = 41\%$  (10%) and  $s = 18\%$  (4%). Thus, GS+FGPA is not accurate at all in this context. This fact should be born in mind when dealing with low-resolution DM-only simulation with  $\sim 100$  cMpc boxes. On the contrary, the accuracy is dramatically improved by 3D-IMS. For the PDF, we have  $m = 18\%$  (13%) and  $s = 10\%$  (6%) for GS+FGPA (3D-IMS). Therefore, also this statistics is better reproduced by 3D-IMS. We note that the precision achieved for the flux 1DPS and PDF is of the same order of what we obtained when we extracted the DM density field from the same hydrodynamic simulation used for the calibration.

In Figure 2.13 we plot the 3DPS of the flux given by the hydrodynamic simulation and obtained with GS+FGPA and 3D-IMS, applied to the DM-only simulation. The color coding is the same as in Figure 2.12. The different panels show the 3DPS in four evenly spaced  $\mu$ -bins. From our definition of  $\mu$ , the bin  $0.0 < \mu < 0.25$  corresponds to modes farther from the line-of-sight, whereas the bin  $0.75 < \mu < 1.0$  is the closest to it. Below the plots obtained for each bin, we show the relative difference of the results of each method with respect to the ones given by L80N4096. Once again, we applied the same analysis technique adopted throughout this work, obtaining that  $m \approx 10\%$  for 3D-IMS in all  $\mu$ -bins. On the contrary, the accuracy of GS+FGPA is strongly dependent on the  $\mu$ -bin considered. We have  $m = 10\%$  in the  $\mu$ -bin farthest from the line-of-sight, whereas  $m = 58\%$  in the bin closest to it. These results are consistent to the findings presented in the previous sections.

The accuracy of 3D-IMS in reproducing the 1DPS, 3DPS and PDF of the flux in redshift space is higher than in the case of GS+FGPA. The accuracy is comparable to the results obtained when applying 3D-IMS to the DM field extracted from our reference simulation. To probe the limitations of our technique, we applied the analysis explained in the present section also to the  $256^3$  Gadget run, adopting the same calibrating simulation. We verified that the ratio of the accuracy of the two methods does not change significantly. In conclusion, our method is solid when applied to a large-box low-resolution DM-only simulation. This achievement makes our method attractive for studies requiring both large volumes and high resolution simulations, for which running hydrodynamic simulations is not a viable option. One example is modeling the signature of the BAO on the  $\text{Ly}\alpha$  forest flux power spectrum.

The results presented in this section clearly show that GS+FGPA can yield a very poor accuracy with respect to what would be obtained through a hydrodynamic simulation. Any result claimed after applying this technique should then be considered carefully. Future works making use of GS+FGPA should refer to Figures 2.9 and 2.10 to assess the error intrinsic in the method adopted.

1D-IMS has not been validated for a large-box DM-only run because of the lack of a recipe to model the flux 1DPS (e.g. an analytic fitting formula), extending it to large

scales. Though, we are confident that in future work such fitting procedure could be provided. If such technique becomes available, we do not expect that 1D-IMS will fail the test presented in this section. Indeed, 1D-IMS and 3D-IMS are both grounded on the Iteratively Matched Statistics technique. We proved that 3D-IMS is accurate when applied to a large-box DM-only run, so this is encouraging for 1D-IMS as well.

## 2.6 Comparison to Previous Works

With our analysis technique, we defined a criterion to assess the accuracy in reproducing 1DPS, PDF and 3DPS of flux skewers, which we wish will be used also by other authors in the future. This would make the comparison with upcoming works more direct and straightforward. It is of course interesting to compare our results with previous work. We do that at the best of our possibilities, since the statistics discussed in the relevant literature are not always the same as the ones considered by us. Furthermore, it is the first time that the performances of approximate methods in reproducing the 3DPS of flux are quantified.

Gnedin & Hui (1998) proposed hydro-particle mesh (HPM), a method to describe baryonic pressure as a modification of the gravitational potential in collisionless simulations. They compare the results of their own technique with two reference hydrodynamic simulations. After computing 300 flux skewers at  $z = 3$ , they found out that the mean error on the fractional flux decrement is smaller than 10% in the whole dynamic range. They do not consider other statistics of the flux, but they show that the accuracy in reproducing the column density distribution is around 13%. They claim that HPM would be suitable when an accuracy of 10-15% is needed in the modeling. Both our methods and GS+FGPA (with appropriate smoothing length) result in higher accuracy.

Meiksin & White (2001) also used the HPM technique to compute the flux field. In addition, they considered an N-body particle mesh code, from which they computed the flux using GS+FGPA. They show that the two methods yield the same cumulative distribution of the flux within  $\sim 10\%$  accuracy, for four different cosmological models. The cumulative distributions of column density and Doppler parameter differ up to  $\sim 10\%$  and  $\sim 20\%$ , respectively. Although we consider the PDF and not the cumulative PDF of the flux, their findings agree with our results for the PDF given by GS+FGPA.

Viel et al. (2002) tested GS+FGPA against a smoothed particle hydrodynamic simulation. Furthermore, they developed a hydro-calibrated approximate method to predict the Ly $\alpha$  forest, based on an adaptive filtering scale for the DM density. Although the accuracy of these techniques in recovering the logarithm of the flux PDF given by the hydrodynamic simulation is not quantified, it can be inferred from their plots that GS+FGPA recovers such statistics on average better than 10%, even though the agreement looks worse in certain regions of flux (e.g. around 0.2 or 0.8). The PDF is

reproduced much better by the hydro-calibrated method. Its accuracy can be estimated through eye-balling to be at percent level. It would then mean that 3D-IMS is comparable to the method provided by Viel et al. (2002), as far as the flux PDF is concerned. 1D-IMS still performs much better, matching this statistics by construction.

Font-Ribera et al. (2012a) introduced a method to generate mock data sets for the measurement of the Ly $\alpha$  forest flux correlations at large scales. As a first step, a Gaussian random field is generated along a set of line of sights. Flux skewers are then generated transforming such a field via an exponential function dependent on two parameters, which are tuned to reproduce a certain flux PDF. The flux power spectrum can be modeled to the desired shape by choosing the power spectrum of the initial Gaussian random field accordingly. Bautista et al. (2015) compared the mock data obtained through this technique with real high-redshift quasar spectra from BOSS. It can be deduced from the plots presented that the agreement between the real and mock 1D correlation functions is almost perfect for small wavelength separations, but it can get as worse as  $\sim 30\%$  for larger separations. The mock and measured 3D correlation functions are compatible within  $\pm 1\sigma$ . The advantage of the described method is that one needs modeling neither a large-box density field nor a velocity field, but only generating the required sample of mock skewers. On the other hand, IMS has proved to be accurate at reproducing the flux PDF, 1DPS, and 3DPS even on scales much larger than the ones probed by the calibrating hydrodynamic simulation.

LyMAS method (Peirani et al., 2014) also matches the flux PDF by construction. In its simplest version, this method consists of two hydro-calibrated transformations of the matter density field. Qualitatively, it can be seen that the method reproduces well the 1DPS given by the reference hydrodynamic simulation, except for  $k \gtrsim 1.5/c\text{Mpc}$ . However, LyMAS can be extended with two further transformations of the flux field (LyMAS full scheme). In this way, the accuracy of the 1DPS is dramatically improved, although this was not quantified in Peirani et al. (2014). Flux skewers appear reasonable only in the full scheme, while in the simplest incarnation they are quite noisy.

Comparing LyMAS to our methods, it certainly does better than 3D-IMS at reproducing the PDF. In this respect, it is as good as 1D-IMS, since both match the flux PDF by construction. It also looks like LyMAS recovers the 1DPS given by the hydrodynamic simulation to a very high accuracy. Likewise, 1D-IMS reproduces the 1DPS almost perfectly and accurately reproduces the 3DPS as well. Furthermore, both 3D-IMS and 1D-IMS provide good-looking skewers applying simple transformations. The LyMAS methodology may be improved by also using the velocity field of the N-body run, which is currently being neglected.

An important feature of 3D-IMS is that it matches the flux provided by the hydrodynamic simulation in real space. This sets the correct filtering scale, allowing us to explore many values of  $T_0$  and  $\gamma$  when computing the redshift-space flux. On the contrary, LyMAS connects the dark matter density directly with the redshift-space flux, so

each choice of  $(T_0, \gamma, \lambda_G)$  requires an additional hydrodynamic simulation. Therefore, in this regard, 3D-IMS appears to be more flexible than LyMAS.

Recently, Lochhaas et al. (2015) used LyMAS to predict the cross-correlation between DM halos and Ly $\alpha$  forest flux, and compared it to quasars-damped Ly $\alpha$  systems cross-correlation measurements from BOSS (Font-Ribera et al., 2013, 2012b). From the plots presented, one can tell that the DM halos-Ly $\alpha$  forest cross-correlation given by LyMAS reproduces very well the results of their calibrating hydrodynamic simulation (Horizon-AGN; Dubois et al. 2014). However, other statistics relevant for our work, like the 1DPS, are not computed.

## 2.7 Chapter Summary

In this Chapter, we investigated approximate methods to obtain statistical properties of Ly $\alpha$  forest from N-body simulations. We focus our attention on the PDF, 1DPS and 3DPS of the flux field, comparing results of approximate methods with a reference hydrodynamic simulation.

We studied the limitations of the FGPA, which is the basis of many approximate methods. The primary sources of error are the differences between DM and baryon velocity fields and, to a smaller degree, the impact of scatter in the temperature-density relationship of the IGM. The accuracy of the FGPA in reproducing the 1DPS and PDF is around 2%, and around 5% for the 3DPS. We also assessed the accuracy of the widely used Gaussian smoothing technique, combined with the FGPA (GS+FGPA). This method consists in mocking the baryon density by smoothing the matter density with a Gaussian kernel. Such field is then used to compute the flux within the FGPA. The accuracy at which the statistics of the flux given by the reference hydrodynamic simulation is reproduced varies a lot with the choice of the smoothing scale  $\lambda_G$ . We explored a wide range of smoothing lengths and found out that the best accuracy achieved for 1DPS and 3DPS is  $\sim 7\%$  and  $\sim 5\%$ , respectively (at  $\lambda_G = 57$  ckpc), and  $\sim 4\%$  ( $\lambda_G = 114$  ckpc) for the PDF. For smoothing scales  $\gtrsim 171$  ckpc the mean accuracy is worse than 20%. This dependence of GS+FGPA on the smoothing scale is rather unfortunate, as the “optimal” smoothing scale is guaranteed to differ for models with different thermal IGM history. As one does not know a priori this optimal value, in practice it means that works using *any* particular smoothing scale will have errors varying in an uncontrolled manner.

To remedy these problems, we have developed two new methods, 3D-IMS and 1D-IMS, based on the idea of Iteratively Matched Statistics (IMS). Their starting point is also Gaussian-smoothing the matter density on a certain scale, which corresponds to the mean interparticle spacing of the simulation considered. In 3D-IMS, smoothing is followed by matching the 3D power spectrum and PDF of the flux in real space to the reference hydrodynamic simulation. With 1D-IMS, we additionally match the 1DPS

and PDF of the flux in redshift space. In contrast to GS+FGPA, 3D-IMS is much less dependent on the initial smoothing length. It reproduces the 3D power spectrum of the flux in redshift space as accurate as GS+FGPA when smoothing scales are small, but performs significantly better for large smoothing scales, with an accuracy of  $\sim 7\%$  even for smoothing scales as large as  $\sim 500$  ckpc. This is a very important property, as it brings significantly more accurate models of the Ly $\alpha$  forest statistics in large-volume simulations where the mean interparticle spacing has to be large due to computational constraints. The 1D-IMS method matches flux 1DPS and PDF by construction. It still performs equally well, or better than GS+FGPA in reproducing the 3DPS ( $\sim 5\%$ ). It is not necessary to use both methods; one can use 3D-IMS only, reproducing the flux 3D power spectrum accurately, at the expense of a lower accuracy for 1DPS and PDF.

These assessments stand for modeling the Ly $\alpha$  forest even in high resolution N-body simulations, but are especially prominent when large-volume (thus coarse resolution) N-body simulations are used. We have showed that IMS approximate methods significantly outperform GS+FGPA in such case. Indeed, through the iterative procedure, our method correctly recovers small-scale physics which is otherwise not present in low-resolution simulations.

We applied 3D-IMS on an 114 cMpc N-body simulation, a 14 Mpc hydrodynamic simulation as calibration. We then compared the results of 3D-IMS with the predictions of an 114 cMpc high-resolution hydrodynamic simulation, verifying that the PDF, 1DPS and 3DPS are matched within 10%, 10% and 13%, respectively. By contrast, GS+FGPA reproduces the same statistics given by the reference 114 Mpc hydrodynamic simulation within 18%, 41% and 58%. Therefore, 3D-IMS improves the accuracy with which GS+FGPA predicts the 1DPS and 3DPS by a factor of 4. In addition, 3D-IMS appears more robust and easy to implement, constituting an improvement over previous techniques. We reiterate that the accuracy estimated for 3D-IMS should be considered as a lower limit, because we calibrated the method with the 114 cMpc hydrodynamic simulation, which is converged resolution-wise, but not in volume. Thus, repeating the analysis in section § 2.5 with a fully converged calibrating hydrodynamic simulation would probably yield an even higher accuracy for 3D-IMS.

With the fitting formula for the 3D real-flux power spectrum provided by Kulkarni et al. (2015), it is particularly straightforward to “paint” such statistics on a larger box (see discussion in § 2.3.1 and § 2.5). It will be useful to achieve an analogous formula for the 1DPS of the redshift-space Ly $\alpha$  flux, as that would allow to straightforwardly apply 1D-IMS on larger boxes, in the same fashion adopted for 3D-IMS in § 2.5.

Our methods have applicability in any context where large-box simulations are needed. The high accuracy of 3D-IMS and 1D-IMS at large smoothing lengths demonstrates that the hydro-calibrated mappings are able to “paste” information about the small-scale physics of the IGM not present in a large volume simulation, without compromising large-scale statistics. To be quantitative, we showed that at  $\lambda_G = 228$  ckpc 3D-IMS

matches predicts several Ly $\alpha$  flux statistics within  $\sim 10\%$ . Since in any realistic situation  $\lambda_G$  has to be at least as large as the inter-particle separation, it means that one would achieve the aforementioned accuracy applying 3D-IMS to a collisionless simulation with a trillion particles in a  $\sim 2.5$  cGpc box. This size is large enough to comfortably study the signature of the BAO signal in the Ly $\alpha$  forest. As a reference, in context of BOSS survey White et al. (2010) ran a suite of N-body simulations with a box size of 1.02 cGpc and  $4000^3$  particles (inter-particle separation 260 ckpc), applying the Gaussian-smoothing technique. Currently, the state of the art for N-body simulations is represented by the “Millennium XXL” (box size 4.11 cGpc,  $6720^3$  particles; Angulo et al. 2012), “Outer Rim” (box size 4.3 cGpc,  $10240^3$  particles; Habib et al. 2012, 2013) and “Dark Sky” (box size 11.5 cGpc,  $10240^3$  particles; Skillman et al. 2014) simulations.

The BAO signal can be modulated by UV background fluctuations, which are coupled to fluctuations in the mean free path of ionizing photons on large scales (Gontcho A Gontcho et al., 2014, Pontzen, 2014). For a proper modeling, one needs a simulation with a box size much larger than the mean free path (Davies & Furlanetto, 2015), which is of order the BAO scale at  $z \sim 2.5$  (Worseck et al. 2014 and references therein). Therefore, one would have to run radiative transfer simulations with box sizes of the order of 1 cGpc — far beyond current computational capabilities. The high quasar density in the BOSS survey allows measuring the 3D power spectrum, which can be exploited to improve cosmological constrains and/or constrain IGM thermal properties (McQuinn et al., 2011, McQuinn & White, 2011). Finally, our technique could help in modeling the cross-correlation between Ly $\alpha$  forest and HI 21 cm signal (Guha Sarkar & Datta, 2015), as well as between CMB lensing and Ly $\alpha$  forest (Vallinotto et al., 2009, 2011).

## Chapter 3

# Constraining Simulations with $\text{Ly}\alpha$ Absorption Profiles around Galaxies

As anticipated in Chapter 1, the CGM of galaxies and QSOs can be probed via absorption line measurements in the spectra of background QSOs passing at small transverse separations from the foreground object. For example, the spectra of 15 highly luminous QSOs in the Keck Baryonic Structure Survey (KBSS) revealed that the CGM of foreground star-forming galaxies in the redshift range  $2 \lesssim z \lesssim 3$  presents an excess of neutral hydrogen (HI) absorption with respect to the IGM (Rakic et al., 2012, Rudie et al., 2013, 2012b). Turner et al. (2014) confirmed these results and also observed an increased optical depth for some metal lines (CIII, CIV, NV, OVI, SiIV) in the vicinity of the galaxies. Other observations probed the  $\text{Ly}\alpha$  transmission up to galactocentric distances of  $\sim 10 h^{-1} \text{Mpc}$ , thus providing insight into the physics of both the CGM and IGM (Adelberger et al., 2005a, 2003, Crighton et al., 2011).

The Quasars-Probing-Quasars (QPQ) project uncovered a large sample of projected QSO pairs with small impact parameters (Hennawi, 2004, Hennawi et al., 2010, 2006b), which enabled the first studies of the CGM of QSOs (Hennawi & Prochaska, 2007, Hennawi et al., 2006a). Using an enlarged statistical sample, Prochaska et al. (2013)

---

I am the main contributor of the work presented in this Chapter. I developed the semi-analytic method presented in § 3.2.4 and conducted the analyses in § 3.3 and § 3.4 entirely by myself. I developed several ideas and I conducted the analysis in § 3.2, writing several codes myself and modifying some existing codes written by Prof. Dr. Joseph F. Hennawi. The conceptual foundation of the project is his idea, and he provided guidance, suggestions and new ideas throughout the project. Dr. Jose Oñorbe contributed with ideas and advice, and helped with the coding necessary for the analyses. Dr. Zarija Lukić ran the Nyx simulation adopted in the project. This work is still unpublished (Sorini et al., 2017, to be submitted). We are thankful to Prof. Dr. Volker Springel for sharing a piece of code, which we modified, to bin the outputs of the Illustris simulation into a regular Cartesian grid. We thank Andreu Font-Ribera for providing us with the table of the BOSS data used in this work.

observed an excess of Ly $\alpha$  absorption within 1 Mpc from the foreground QSOs (see also Hennawi et al. 2006a and Hennawi & Prochaska 2007). This enhanced absorption due to HI, as well as metals (Lau et al., 2016, Prochaska et al., 2014), implies the presence of a substantial reservoir of cool ( $T \sim 10^4$  K) metal-enriched gas around quasars.

Background QSOs can be utilized to study Damped Ly $\alpha$  Absorbers (DLAs) as well. Font-Ribera et al. (2012b) measured the large-scale Ly $\alpha$ -DLA cross-correlation with data from BOSS (Ahn et al., 2012). More recently, Rubin et al. (2015) exploited close QSO pairs, in which the sightline of one of the two QSOs passed through an intervening DLA, while the other QSO probed Ly $\alpha$  and metal line absorptions at a certain impact parameter from the DLA. Also these measurements indicate an excess in the Ly $\alpha$  absorption within 200 kpc from the DLA.

There is thus now a large amount of data characterizing the strength of HI Ly $\alpha$  absorption in the CGM of galaxies, QSOs, and DLAs. The strength of HI absorption is determined by the abundance of cool ( $T \lesssim 10^5$ ) gas in the CGM, which is in turn determined by the physical state of CGM gas (density, temperature, cloud size). While the inflow of cool material from the IGM to the CGM is predicted ab initio by cosmological hydrodynamic simulations, it is ultimately the interplay between these inflow and complex feedback processes that determine its physical state. Nonetheless, feedback processes are still poorly understood, and their implementations in numerical simulations vary from case to case (Somerville & Davé, 2015, for a review). The feedback prescriptions can be constrained comparing the predictions of simulations with measurements of absorption features in the CGM.

Most of numerical works in the literature have focused primarily on the covering factor of optically thick absorbers. Several zoom-in simulations (Ceverino et al., 2010, 2012, Dekel et al., 2013, Shen et al., 2013) reproduced Rudie et al. (2012a) observations of this quantity within the virial radius of star-forming galaxies, but not the Prochaska et al. (2013) measurements around QSOs (Faucher-Giguère et al., 2015, Fumagalli et al., 2014). The latter observations were well reproduced by Faucher-Giguère et al. (2016), improving the mass resolution of FIRE zoom-in simulations (Hopkins et al., 2014), and by Rahmati et al. (2015) by means of the EAGLE suite of hydrodynamic cosmological simulations (Crain et al., 2015, Schaye et al., 2015). With a set of simulations based on the moving-mesh hydrodynamic code AREPO (Springel, 2010a), Suresh et al. (2015) reproduced the measured covering fraction of DLAs around  $10^{12} M_{\odot}$  halos, within the ample error bars of the observations.

Another well-studied statistics is the column density distribution function (CDDF) of HI absorbers. A good agreement with observations (Kim et al., 2002, Péroux et al., 2005, Prochaska & Wolfe, 2009, Zwaan et al., 2005) was found by Rahmati et al. (2013), post-processing the OWLS suite of cosmological simulations (Schaye et al., 2010) with radiative transfer. The NIHAO suite of zoom-in cosmological simulations (Wang et al.,

2015) reproduces overall well Prochaska et al. (2011) and Tumlinson et al. (2013) measurements of the HI CDDF as a function of the impact parameter from, and luminosity of, the foreground galaxy (Gutcke et al., 2017), although some data points in the lowest luminosity bin are underestimated. The AREPO code is able to recover kinematics and CDDF of DLAs at  $z = 3$ , as well as the total DLA abundance and metallicity in the range  $2 < z < 4$ , with the exception of the DLA abundance at  $z < 3$  (Bird et al. 2015, 2014; see also Bird et al. 2013).

There is another body of numerical works that are mainly concerned with Ly $\alpha$  absorption around foreground LBGs, DLAs or QSOs (Kollmeier et al., 2006, 2003, Meiksin et al., 2017, 2014, 2015, Turner et al., 2017). Such works compare full-box hydrodynamic simulations, equipped with different feedback implementations, with observations of the Ly $\alpha$  flux decrement around LBGs (Adelberger et al., 2005a), the equivalent width around star-forming galaxies (Rakic et al., 2012, Steidel et al., 2010) and QSOs (Prochaska et al., 2013), and the median Ly $\alpha$  optical depth around LBGs (Rakic et al., 2012, Turner et al., 2014). In general, all these works exhibit good agreement with the observations, except for the data points within the virial radius of the foreground objects.

In this work, we follow the line of research comparing simulations and observations of Ly $\alpha$  absorption around QSOs, DLAs and LBGs, with the aim of understanding the physics of the CGM and constraining feedback prescriptions. On top of observations considered in previous works (Adelberger et al., 2005a, 2003, Crighton et al., 2015, Prochaska et al., 2013, Rubin et al., 2015, Turner et al., 2014), we also examine the measurements of the Ly $\alpha$ -DLA and Ly $\alpha$ -QSO cross-correlation from BOSS (Font-Ribera et al., 2013, 2012b). For the first time, we convert them into an estimate of the mean Ly $\alpha$  transmitted flux at different impact parameters from the foreground object. This allows us to combine large-scale observations with the aforementioned measurements of Ly $\alpha$  absorption at small impact parameter. In this way, we can simultaneously constrain the physics of both CGM and IGM, spanning impact parameters ranging from  $\sim 25$  kpc up to  $\sim 20$  Mpc.

We use two state-of-the-art hydrodynamic cosmological simulations to make our comparison with observations: the publicly available runs of the “Illustris” simulation (Nelson et al., 2015) and a large-box run of the hydrodynamic code “Nyx” (Almgren et al., 2013, Lukić et al., 2015). The former is endowed with both stellar and AGN feedback, whereas the latter has no prescription for star formation, no metals and no feedback. We find out that two simulations match the data outside the virial radius, but underpredict the Ly $\alpha$  absorption within the virial radius of QSOs and LBGs. Furthermore, the simulations predict a different Ly $\alpha$  mean flux out to  $\sim 2$  Mpc away from QSOs and DLAs, well beyond the virial radius. We argue that, out to 3 virial radii from these objects, the differences can be caused by the AGN feedback prescription in Illustris. The mean Ly $\alpha$  flux profile proves then to be an excellent probe of the physics implemented in simulations. As such, matching this observable should become a new fundamental test for feedback prescriptions, complementary to other already widely-used measurements

(e.g. Behroozi et al., 2013, Guo et al., 2011, Moster et al., 2013). We also develop a semi-analytic technique to alter the temperature of the CGM in post-processing, and apply it on our Nyx run. Exploiting this technique, we show that the discrepancy with the data within the virial radius could be mitigated if the simulations predicted a cooler CGM.

This Chapter is organized as follows. In § 3.1, we describe the details of the simulations adopted. In § 3.2, we explain how we constructed the sample of simulated spectra of Ly $\alpha$  absorption. We compare the predictions of the simulations with observations in § 3.3. In § 3.4, we discuss the physical implications of our results, as well as the impact of possible systematic errors in our analysis. We summarize our results and discuss the perspectives of this work in § 3.5.

## 3.1 Simulations

To study the Ly $\alpha$  absorption in the CGM, we used a large-box run of the Nyx hydrodynamic code, and the publicly available Illustris cosmological hydrodynamic simulation. In this section, we briefly describe the main characteristics of the two simulations.

### 3.1.1 Nyx

In this work, we adopted a different run of the Nyx (Almgren et al., 2013, Lukić et al., 2015) hydrodynamic code, already introduced in § 2.1. Nyx treats dark matter as self-gravitating Lagrangian particles, and modeling baryons as an ideal gas on a uniform Cartesian grid. Shock waves are accurately captured by solving the Eulerian equations of gas dynamics through a second-order-accurate piecewise parabolic method. The adaptive mesh refinement (AMR) option provided by Nyx is switched off in the present work, since we are interested in simulating the Ly $\alpha$  absorption in the entire box, and not only within the most overdense regions (i.e. halos). More details about the numerical methods implemented in Nyx and relevant scaling tests can be found in Almgren et al. (2013).

The physical processes necessary to model the Ly $\alpha$  forest are included in the code. The gas is considered to have a primordial composition, with abundances  $X_p = 0.76$  and  $Y_p = 0.24$ , respectively. Furthermore, the inverse-Compton cooling off the microwave background is included, and the net loss of the thermal energy resulting from atomic collisional processes is accounted for. The values of the recombination, collisional ionization, dielectric recombination rates, as well as cooling rates, which are used in this work, are given in Lukić et al. (2015). The ionizing ultraviolet background (UVB) is given by Haardt & Madau (2012). Self-shielding is modeled following Rahmati et al. 2013 (see § 3.2.3 for further discussion). Nyx does not incorporate any star formation or

feedback prescription. Since there is no star particle, when we compute the Ly $\alpha$  optical depth in a certain cell, we threshold the density in that cell to  $\delta_{\text{th}} = 1000$ <sup>1</sup>.

The initial conditions are generated through the MUSIC code (Hahn & Abel, 2011) and a CAMB (Howlett et al., 2012, Lewis et al., 2000) transfer function. All simulations are initialized at redshift  $z_{\text{ini}} = 159$ , making sure that nonlinear evolution is not compromised (for a detailed discussion on this issue see, e.g., Oñorbe et al., 2014). The cosmological model assumed is the  $\Lambda$ CDM model with parameters consistent with Planck data (Planck Collaboration et al., 2016):  $\Omega_{\text{m}} = 0.3$ ,  $\Omega_{\Lambda} = 1 - \Omega_{\text{m}} = 0.7$ ,  $\Omega_{\text{b}} = 0.047$ ,  $h = 0.685$ ,  $\sigma_8 = 0.8$ .

In this work, we consider the snapshots at redshift  $z = 2.4$  and  $z = 3$  of a run with a volume of 146 cMpc,  $4096^3$  gas cells and as many DM particles. This run represents a state-of-the-art hydrodynamic cosmological simulation. The box size is comparable with the largest hydrodynamic cosmological simulations in the literature (e.g. Crain et al., 2017, Vogelsberger et al., 2014b). The resolution of 35.6 ckpc for baryons grants a precision at percent level in the PDF of the Ly $\alpha$  forest flux and within  $\sim 5\%$  in the 1D Ly $\alpha$  flux power spectrum up to  $k \sim 0.1 \text{ s km}^{-1}$  (Lukić et al., 2015).

### 3.1.2 Illustris

Illustris (Genel et al., 2014, Sijacki et al., 2015, Vogelsberger et al., 2014a,b) is a hydrodynamic simulation based on the AREPO code (Springel, 2010a). Dark matter is treated with a Lagrangian approach, while baryons are described as an ideal gas on a moving mesh constructed with a Voronoi tessellation (Springel, 2011). A Tree-PM scheme (Xu, 1995) is used to compute gravitational forces; a particle-mesh method calculates long-range forces, while a hierarchical algorithm (Barnes & Hut, 1986) is used to determine short-range forces. Gas dynamics is implemented by solving the viscosity-free Euler equations.

Several fundamental astrophysical processes for galaxy formation are included: primordial and metal-line cooling, a sub-resolution model of the interstellar medium, stochastic star formation above a density threshold of  $0.13 \text{ cm}^{-3}$ , gas recycling and chemical enrichment. Kinetic stellar feedback is implemented through supernovae-driven winds, determined by the velocity dispersion of the host halo (Vogelsberger et al., 2013). Illustris includes also super-massive black hole seeding, accretion and merging. The relative AGN feedback implementation follows Sijacki et al. (2007), presenting a dual modeling, which depends on the black hole accretion rate: a radiatively efficient model and

<sup>1</sup>We also considered other values of  $\delta_{\text{th}}$ , ranging from 200 to 3000. By visual inspection, we verified that simulating the Ly $\alpha$  absorption spectra with  $\delta_{\text{th}} = 1000$  gives physically reasonable results.

a “radio-mode”. The free parameters governing the feedback prescriptions were constrained matching the overall star formation efficiency predicted by smaller scale simulations (Vogelsberger et al., 2013) to observed data (Behroozi et al., 2013, Guo et al., 2011, Moster et al., 2013).

Photo-ionization and heating are implemented with the UVB by Faucher-Giguère et al. (2009b). Self-shielding and ionization from nearby AGN are taken into account. In analogy to what we are doing with the Nyx simulation, when we compute the Ly $\alpha$  optical depth from Illustris, we also include collisional ionization and self-shielding, following Rahmati et al. 2013 (see § 3.2.3 for further discussion).

The simulation is initialized at  $z_{\text{ini}} = 127$  (see Vogelsberger et al., 2014a, for details). The cosmological model assumed is the  $\Lambda$ CDM model with parameters consistent with the 9-year data release of WMAP (Hinshaw et al., 2013):  $\Omega_{\text{m}} = 0.2726$ ,  $\Omega_{\Lambda} = 1 - \Omega_{\text{m}} = 0.7274$ ,  $\Omega_{\text{b}} = 0.0456$ ,  $h = 0.704$ ,  $\sigma_8 = 0.809$ ,  $n_s = 0.963$ . In this work, we utilize the snapshots at redshift  $z = 2.44$  and  $z = 3.01$ . Unless otherwise stated, we refer to the “Illustris-1” run, which is the highest-resolution one. The volume of the simulation is  $(106.5 \text{ cMpc})^3$ ; with  $1820^3$  DM particles, and as many gas Voronoi cells, the DM mean interparticle separation is 58.5 ckpc. The mass resolution is  $6.3 \times 10^6 M_{\odot}$  and  $1.3 \times 10^6 M_{\odot}$  for DM and gas, respectively.

## 3.2 Modeling

In this section, we explain how we simulate mock absorption spectra (skewers) to reproduce the observations. Specifically, we consider measurements of the Ly $\alpha$  absorption in the CGM of QSOs, DLAs and LBGs. These data are obtained observing the spectra of background QSOs along lines of sight passing at different impact parameters from the foreground objects.

Reproducing the observations from simulations requires three steps. First of all, we need to select a sample of halos representing the foreground objects. Then, we extract a sample of lines of sight at different impact parameters from each halo, and finally we compute mock Ly $\alpha$  absorption spectra along such skewers.

### 3.2.1 Selection of Halos

To begin with, we need to identify DM halos in Nyx and Illustris, and determine their mass; then, we will select samples of halos hosting the foreground QSOs, LBGs and DLAs. For Illustris, we use the publicly available halo catalogs. Halos are identified grouping at least 32 particles by means of a friends-of-friends (FOF) algorithm (Davis et al., 1985), with a linking length equal to 0.2 times the mean particle separation. All other particle types are then associated to these dark matter halos through a second

Observations <sup>a</sup>	$z$ <sup>b</sup>		$\log(M_{\min}/M_{\odot})$ <sup>c</sup>	
	Nyx	Illustris	Nyx	Illustris
Font-Ribera et al. (2013)	2.4	2.44	12.5	12.4
Prochaska et al. (2013)				
Font-Ribera et al. (2012b)	2.4	2.44	11.7	11.6
Turner et al. (2014)				
Rubin et al. (2015)				
Adelberger et al. (2005a, 2003)	3.0	3.01	11.5	11.5
Crighton et al. (2011)				

<sup>a</sup> Observations considered in this work.

<sup>b</sup> Redshift of the snapshot considered to reproduce the observations in column 1.

<sup>c</sup> Mass threshold for the halos selected to reproduce the observations in column 1.

TABLE 3.1: Parameters used in simulations to reproduce observations.

linking stage (Dolag et al., 2009). We consider the mass of each halo to be the total FOF mass assigned to that halo. In Nyx, halos are identified by finding the topological connected components with density above 138 times the mean density.

In principle, the halo masses in the Nyx and Illustris catalogs may not be consistent, being determined with different halo-finding algorithms and mass definitions. To ensure a physically meaningful comparison, we calibrate the halo masses of the two simulations with the same observations of QSO and LBG clustering. Specifically, for each simulation, we determine the minimum halo mass  $M_{\min}$  above which the 3D auto-correlation function of halos matches the same quantity observed for QSOs (White et al., 2012) and LBGs (Bielby et al., 2011). For the QSO sample, we obtain  $M_{\min} = 10^{12.4} M_{\odot}$  and  $M_{\min} = 10^{12.5} M_{\odot}$  at  $z \approx 2.4$  in Illustris and Nyx, respectively. For the LBG sample, the value of  $M_{\min}$  at  $z \approx 2.4$  is  $10^{11.6} M_{\odot}$  and  $10^{11.7} M_{\odot}$  in Illustris and Nyx, respectively. These values are consistent with the typical mass of LBG-hosting halos in the KBSS,  $\sim 10^{12} M_{\odot}$  (Adelberger et al., 2005b, Conroy et al., 2008, Rakic et al., 2013, Trainor & Steidel, 2012, Turner et al., 2014).

Regarding the measurements of the Ly $\alpha$  absorption around DLAs, we considered the same sample of halos selected for the LBGs. This choice is justified by the fact that the typical mass of DLAs estimated by Font-Ribera et al. (2012b) from a large sample of BOSS quasar spectra ( $\sim 10^{12} M_{\odot}$ ) is the same as the one measured for LBGs. Identifying the LBG and DLA samples implies the additional assumption that the DLAs lie along the line-of-sight passing through the center of the LBG-hosting halo.

We summarize all derived values of  $M_{\min}$  in Table 3.1. Further details on the procedure to determine  $M_{\min}$  can be found in the Appendix B.

### 3.2.2 Construction of Samples of Skewers

Once we set  $M_{\min}$ , we are left with a list of halos selected from the halo catalog, corresponding to the selection criterion described in the previous subsection. We then need to extract a sample of skewers around each halo, within a certain range of the impact parameter  $b$ . We always adopt the same impact parameter bins as the observations that we want to reproduce, and draw  $10^4$  skewers for each bin.

We randomly draw the position of the first skewer around the first halo in our list from a uniform distribution in  $\log b$  across the impact parameter bin, and a uniform distribution in the phase angle around the halo. We translate the position of the skewer into a regular Cartesian grid. In the case of Nyx, this is simply the grid used in the simulation to describe the evolution of gas. In the case of Illustris, we treat each Voronoi cell as an SPH (Smoothed Particle Hydrodynamics; see Springel 2010b for a review) particle, and bin it into a regular grid (following Bird et al., 2014). The size of a grid cell corresponds to the mean interparticle separation in Illustris. We verified that a finer grid would not change the conclusions of this work appreciably (see the Appendix C for more details).

We extract the gas density, temperature and line-of-sight velocity along the selected skewer, throughout the simulation box. To draw the second skewer of our sample, we consider the second halo in the list and repeat the aforementioned procedure. We proceed in this way until we reach the last halo of the list. Since the number of skewers is larger than the number of halos, the following skewer is again extracted around the first halo of the list. If the transverse distance between this skewer and the one previously drawn in the same halo is less than 3 grid cells, the coordinates of the new skewer are re-drawn. This check is made any time a new skewer has to be drawn around a halo which has previously been considered to extract other skewers. With the procedure adopted, we minimize the number of skewers per halo and set a minimum transverse distance among skewers around the same halo. In this way, we avoid having two highly correlated, or even identical, skewers.

### 3.2.3 Simulating HI Absorption

Once all density, temperature and velocity skewers are extracted, we can compute the Ly $\alpha$  absorption spectra. To do this, we first of all need to determine the HI density  $n_{\text{HI}}$  in each pixel.

The total hydrogen density in each cell is simply obtained multiplying the gas overdensity by the cosmic fraction of hydrogen. The ionized fraction of hydrogen is then set by the balance between recombination and all ionization processes, which depend on the density and temperature of the gas. Performing radiative transfer on top of the EAGLE cosmological hydrodynamic simulations, Rahmati et al. (2013) showed that the dependence of the photoionization rate on the hydrogen density  $n_{\text{H}}$  can be divided

into three regimes, corresponding to different dominant ionization mechanisms: photoionization and collisional ionization below and above the self-shielding density threshold ( $\sim 6.0 \times 10^{-3} \text{ cm}^{-3}$  at  $2 < z < 3$ ), respectively, and recombination radiation for  $n_{\text{H}} \gtrsim 10^2 \text{ cm}^{-3}$ . Rahmati et al. (2013) provided an analytic fit to the photoionization rate - hydrogen density relationship, and proved that it is insensitive to the box size and resolution of the simulation, suggesting that it can be used to determine  $n_{\text{HI}}$  in simulations on top of which no radiative transfer has been performed. We then use their parameterization of the photoionization rate - hydrogen density relationship to determine  $n_{\text{HI}}$  in Nyx and Illustris.

After computing  $n_{\text{HI}}$ , we determine the Ly $\alpha$  optical depth  $\tau$  taking into account the thermal broadening of the absorption lines and redshift space distortions (see e.g. Lukić et al., 2015). The Ly $\alpha$  flux is then simply given by  $F = \exp(-\tau)$ . Following the standard approach, we chose the value of the UVB such that the flux skewers are consistent with the observed mean flux of the IGM at the redshift of the observations. This is ensured by randomly drawing  $10^5$  skewers from Illustris and Nyx, and tuning the UVB in both simulations such that the mean flux matches the observations by (Becker et al., 2013). The obtained values of the UVB are then used to compute the transmitted flux along the samples of skewers generated as explained in § 3.2.1 and § 3.2.2.

### 3.2.4 Altering the Temperature of the CGM

To get insight into the connection between HI absorption and temperature of CGM gas, we developed a semi-analytic technique to “paint” different temperature-density relationships in the CGM of Nyx halos. Our goal is to come up with a simple method to alter the temperature of the CGM, depending on a small number of intuitive parameters, and capable of producing physically reasonable skewers.

We begin by visually inspecting hydrogen density and temperature maps around halos. In Figure 3.1 we show these quantities within a one-pixel-thick slice centered around a sample of 4 halos in Nyx (first two columns from the left) and Illustris (last two columns from the left). Every row displays one halo from each simulation, chosen such that their masses agree within 0.2%. The mass of the halos increases from top to bottom. The mass of the halos in the first two rows is typical of LBGs, whereas the remaining rows contain halos with a mass characteristic of QSOs. The side of each slice shown is equal to 5 times the virial radius of the halo at its center. The virial radius  $r_{\text{vir}}$ , marked with a black circle at the center of slice displayed in Figure 3.1, was computed using the Barak Python package<sup>2</sup>.

Let us now focus on the panels of Figure 3.1 relative to Nyx. We notice that, for each halo, the temperature of the CGM broadly traces the underlying hydrogen density in

<sup>2</sup><http://nhmc.github.io/Barak/index.html>

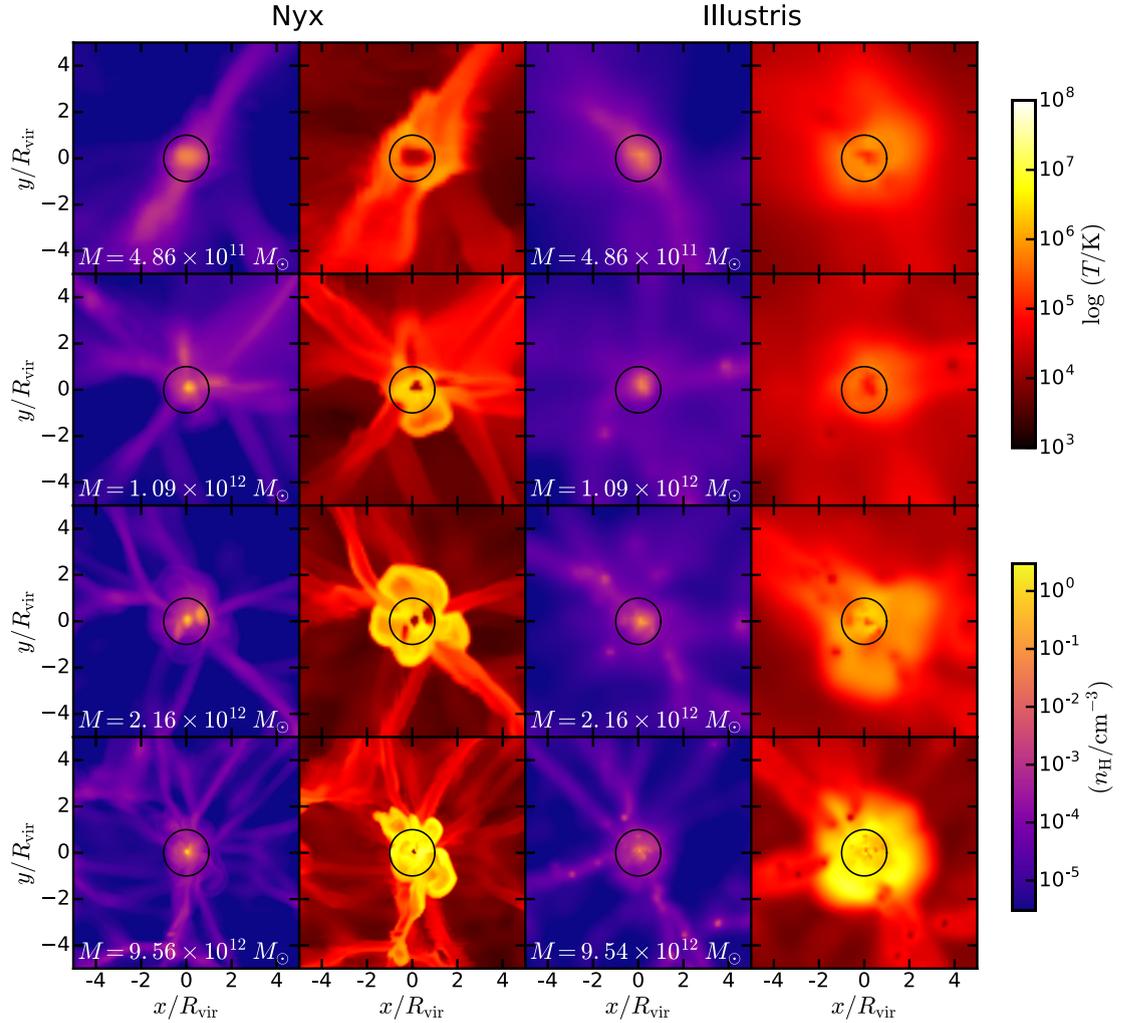


FIGURE 3.1: *Top panels:* Temperature and hydrogen density slices around a sample of 4 halos selected from Nyx and Illustris at  $z = 2.4$  and  $z = 2.44$ , respectively. The first and second columns from the left show the temperature and hydrogen density around Nyx halos, respectively, while the third and fourth columns the same quantities for Illustris halos. The slices are one pixel thick, are centered around the halos, and span an area of  $5 \times 5$  virial radii. The virial radius is marked with a black circle. The mass of each halo is written in the hydrogen density panel. For the most massive halos, Illustris generally presents more extended bubbles of hot gas with respect to Nyx.

low-density regions, as expected given that the IGM follows a tight temperature-density relationship (Hui & Gnedin, 1997). On the other hand, gas cells closer to the center of the halo seem to be generally hotter than those farther out<sup>3</sup>.

Following the hints given by the visual inspection of the hydrogen density and temperature slices, we assume that the temperature of the CGM at a certain point in the halo can, to a first approximation, be modeled as a function  $T(n_{\text{H}}, r)$  of the local hydrogen density  $n_{\text{H}}$  and of the distance from the center of the halo  $r$  only. In general, we do not

<sup>3</sup>The only exception is the tiny central dense and overcool region. This is caused by the fact that Nyx does not convert dense gas into stars, hence gas can evolve to very high density, and for the densest cells cooling becomes more efficient.

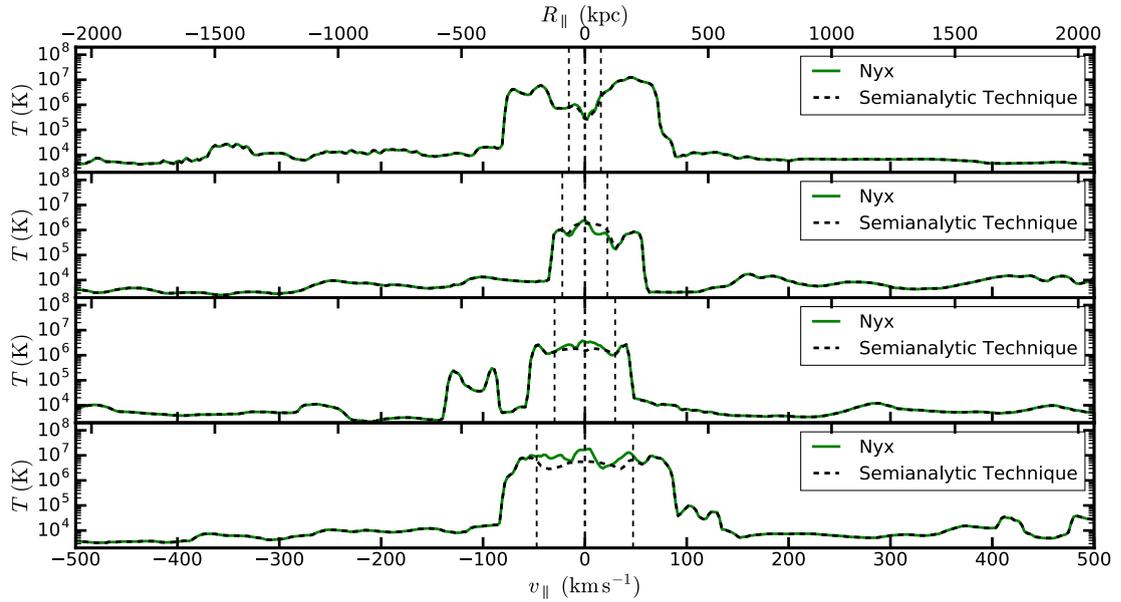


FIGURE 3.2: *Top panels:* A sample of 4 temperature skewers given by Nyx (green lines) and obtained through our semi-analytic technique (black dashed lines). All skewers are taken at 50 kpc impact parameter from one of the halos shown in Figure 3.1, ordered by increasing mass from the top to the bottom panel. All skewers are plotted such that the position of the halo lies at the center of the panel. In all panels, the vertical dashed black lines mark the boundaries of the virial radius. By construction, our semi-analytic technique matches Nyx outside the virial radius. Outside the virial radius, it presents an overall good agreement with Nyx (see the main text for details).

expect this function to have a trivial shape. In order to determine a physically sensible shape for  $T(n_{\text{H}}, r)$ , we first of all select a sample of halos as explained in § 3.2.1 and divide the gas cells into equally spaced logarithmic bins of  $n_{\text{H}}$ . Within each density bin, we compute the mean logarithmic temperature given by Nyx ( $\langle \log T[\text{K}] \rangle$ ) within different radial distance bins, and then fit the resulting mean temperature-radial distance profile with the following function:

$$\langle \log(T[\text{K}]) \rangle = a(n_{\text{H}}) + b(n_{\text{H}}) \log \left( \frac{r}{r_{\text{vir}}} \right), \quad (3.1)$$

where  $a(n_{\text{H}})$  and  $b(n_{\text{H}})$  are the parameters of the fit, which are determined for each  $n_{\text{H}}$  bin.

We used Equation (3.1) to define the aforementioned function  $T(n_{\text{H}}, r)$ , and used such function to re-compute the temperature of the gas within the virial radius of halos in Nyx, leaving the temperature given by Nyx outside the virial radius unmodified<sup>4</sup>.

<sup>4</sup>The transition between the temperature given by Nyx outside the virial radius and the temperature computed through our semi-analytic technique within the virial radius is actually modulated with a smooth approximation of the step function. Given the specific shape of this function, the temperature of the skewers in the Nyx Hot and Nyx Cold models can slightly differ from the values of the pure Nyx run as far as  $\sim 1.2 r_{\text{vir}}$ .

We test the validity of our approach comparing the resulting temperature skewers with the ones given by the unmodified Nyx run. In Figure 3.2, we show the temperature along 4 skewers extracted from Nyx. The skewer depicted in each panel passes at an impact parameter of 50 kpc from one of the halos shown in Figure 3.1, in increasing order of mass from top to bottom. The solid green lines refer to the temperature given by Nyx, while the dashed black lines indicate the temperature re-computed through our semi-analytic technique. We plot only the temperature within a velocity window of  $\pm 500 \text{ km s}^{-1}$  around the position of the halo, which is aligned at the center of every panel and marked with a vertical dashed black line. The other two vertical black dashed lines indicate the boundaries of the virial radius around the position of the halo. The discontinuities in the temperature skewers are due to the virial shocks. Outside the virial radius, our method matches Nyx by construction. Within this range of distance, the agreement is better for lower-mass halos. For the most massive halo ( $\sim 10^{13} M_{\text{odot}}$ , bottom panel) our approach reproduces Nyx within a factor of  $\sim 2$ . The overall good agreement between Nyx and our approach proves that our method is a reasonable and simple technique to semi-analytically describe the temperature of the CGM.

Having verified the robustness of our semi-analytic approach, we now use it to construct two different models for the CGM: one with an overall hotter CGM than predicted by Nyx (“Nyx Hot”), and one with a colder CGM (“Nyx Cold”). We define these models by simply adding a constant to  $a(n_{\text{H}})$  in Equation (3.1), while keeping  $b(n_{\text{H}})$  unchanged. Specifically, in the Nyx Hot model, the additive constant is 1, so that the temperature of the CGM within the virial radius is generally 1 dex higher than the predictions given by Nyx. Differences up to 1 dex in the temperature profiles around halos are also found in other cosmological simulations when comparing runs including feedback with feedback-free runs (e.g. Kollmeier et al., 2006, Nelson et al., 2016, Stinson et al., 2010, 2012, Woods et al., 2014). Hence, although our hot painted model on top of the feedback-free Nyx run is by no means equivalent to a simulation endowed with feedback prescriptions, it mimics reasonably well the heating caused by AGN feedback.

The Nyx Cold model is defined by replacing  $a(n_{\text{H}})$  with  $a(n_{\text{H}}) - 1$ , in analogy with the Nyx Hot model. The temperature of the CGM within one virial radius is then generally one order of magnitude lower than in Nyx. Physically, this model has a less straightforward interpretation. It could simply represent a CGM which cools down more efficiently than expected from current simulations. There may also be further interpretations linked to the physics which is unresolved in the simulations. We shall discuss it in detail in § 3.4. At the current stage, the Nyx Cold model should be considered simply as a means to explore what the Ly $\alpha$  absorption around galaxies would look like if their CGM were colder than what predicted by Nyx.

## 3.3 Results

### 3.3.1 Simulated Spectra

Before computing any statistics of the Ly $\alpha$  absorption lines, we visually inspect a sample of skewers extracted from Illustris, pure Nyx, and Nyx with modified temperature. In Figure 3.3, the top set of five panels shows different physical quantities along the same skewer drawn from the Nyx run, at 50 kpc from the halo shown in the third row of Figure 3.1. From top to bottom, we plot the gas overdensity with the corresponding number density of hydrogen, the gas temperature with the corresponding Doppler parameter, the HI column density, the gas velocity field and the transmitted Ly $\alpha$  flux. The green, blue and red solid lines refer to the pure Nyx run, and to the Cold and Hot Nyx models, respectively. All quantities are centered at the position of the halo along the line of sight, marked with a vertical black dashed line, and plotted within a velocity window of  $\pm 500 \text{ km s}^{-1}$ .

The temperature seems to trace the gas overdensity along the skewer, except for the region within  $\sim 240$  kpc from the center of the halo. This region is delimited by a steep increase of the temperature, which is due to the virial shock. By construction, the temperature of the three models is the same far from the halo, and presents increasingly larger differences as the halo is approached. In turn, these differences impact also the HI column density. The gas velocity field is smooth in the IGM regime. The discontinuity seen at the IGM/CGM interface is due to the virial shock. The flux is the most interesting panel, since that is the actual observable that we are interested in. It is remarkable that, even though we are changing the temperature in a small region along the line of sight, we see huge differences in the flux skewers predicted by the different models. This happens because the peculiar velocities of the gas increase the impact of Doppler broadening features farther from the virial radius. In this way, the Nyx Hot model erases most of the absorption features. In contrast, the Nyx Cold model increases the absorption, giving rise to a deeper extended trough.

In the lower set of five panels in Figure 3.3 we plot the same quantities as in the top set of panels, but for a skewer drawn from the Illustris simulation, at 50 kpc from the halo shown in the third row of Figure 3.1. Also in this case, we can distinguish between IGM and CGM regimes thanks to the virial-shock features in the temperature and gas velocity skewers, although the transition is smoother if compared with Nyx. The hot region around the halo is much more extended in the case of Illustris. Its boundaries span a length of  $\sim 870$  kpc, whereas the corresponding region for the skewer drawn from Nyx is  $\sim 470$  kpc. Consequently, the absorption features are very mild in this example skewer from Illustris.

In general, it seems that the hot component of the CGM spreads farther from the center of the halo in Illustris and Nyx. We shall discuss it into detail in § 3.4.1.

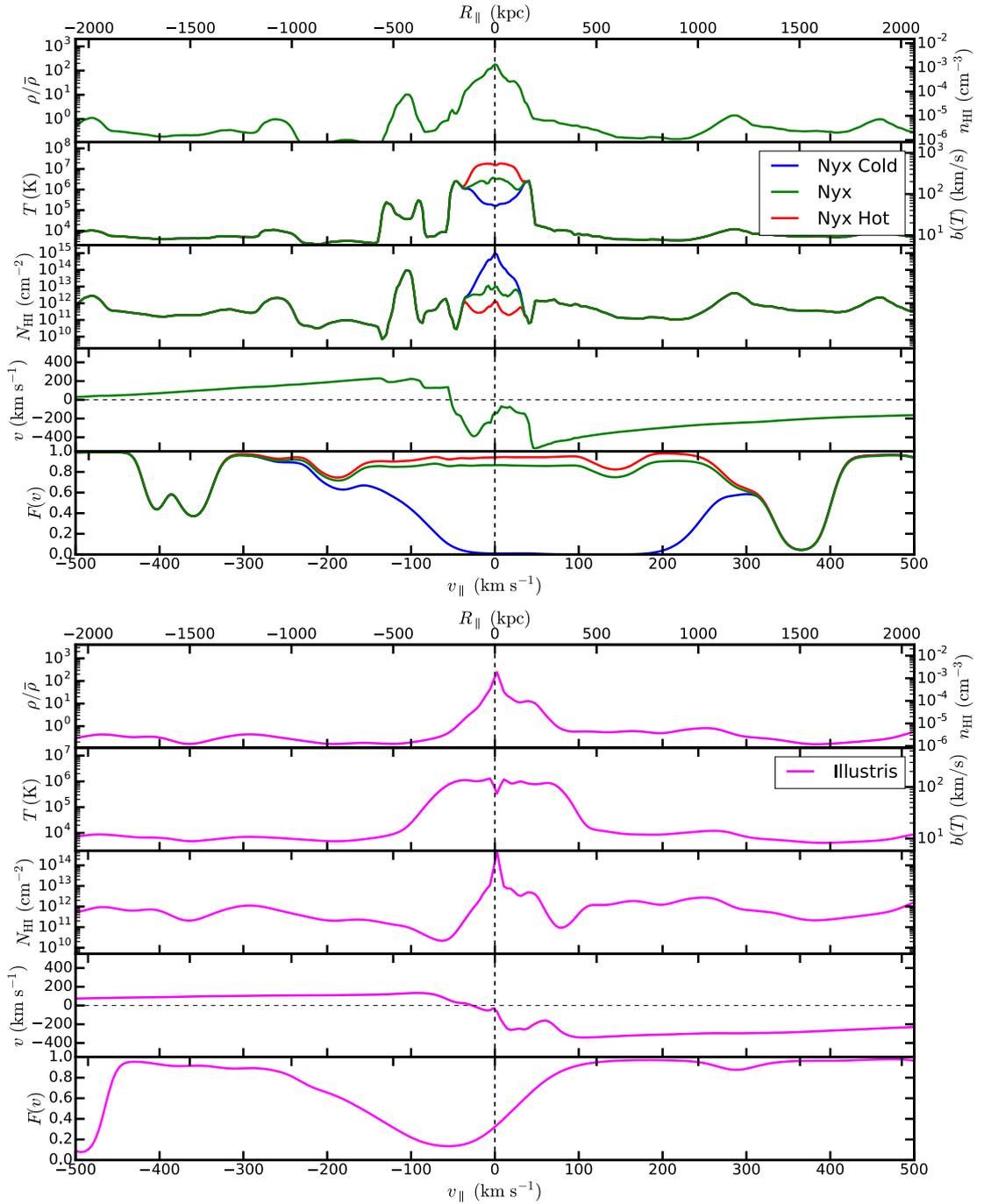


FIGURE 3.3: Different quantities along the same skewer, located at 50 kpc from the Nyx and Illustris halos in the third row from the top of Figure 3.1. In the top set of five panels, from top to bottom, we plot the gas overdensity with the corresponding number density of hydrogen, the gas temperature with the corresponding Doppler parameter, the HI column density, the gas velocity field and the transmitted Ly $\alpha$  flux in a velocity window of  $\pm 500 \text{ km s}^{-1}$  around the Nyx halo. The green solid lines refer to the pure Nyx run. The blue dashed and red dot-dashed lines refer to our Cold Nyx and Hot Nyx models altering the temperature of the CGM in Nyx (see text for details). In the bottom set of five panels, we plot exactly the same quantities as in the top set, this time for the Illustris halo.

### 3.3.2 Comparison with Observations

In this section, we compute various statistics of our sample of flux skewers, and compare them with observations of the Ly $\alpha$  absorption in the CGM of QSOs, DLAs and LBGs.

#### 3.3.2.1 Quasar Hosts

We want to compare the mean flux versus the impact parameter predicted by the models considered, with the observations by Prochaska et al. (2013) and Font-Ribera et al. (2013). The two measurements probe different ranges of impact parameter, complementing each other.

Prochaska et al. (2013) considered a sample of 650 projected QSO pairs in the redshift range  $2 < z < 3$  and with a transverse separation  $< 1$  Mpc. They divided the spectra into 5 bins of impact parameter and, for each of such bins, measured the mean flux within a line-of-sight velocity window of  $\pm 1000 \text{ km s}^{-1}$  around the position of the foreground object. Thus, they determined the mean Ly $\alpha$  flux fluctuations

$$\delta_F = 1 - \frac{\langle F \rangle}{\langle F \rangle_{\text{IGM}}}, \quad (3.2)$$

as a function of the impact parameter. In (3.2),  $\langle F \rangle$  is the mean flux in each impact parameter bin, averaged over all spectra, and  $\langle F \rangle_{\text{IGM}}$  is the mean flux of the IGM at the median redshift of the observations ( $z \approx 2.4$ ). We report these measurements in Figure 3.4 as black squares. The vertical error bars are the  $1\sigma$  errors of the measurements, whereas the horizontal bars show the bin widths.

Font-Ribera et al. (2013) considered a sample of  $\sim 6 \times 10^4$  QSO spectra from BOSS. They measured the QSO-Ly $\alpha$  cross-correlation function in different bins of transverse and line-of-sight separation. This is actually equivalent to measuring the mean Ly $\alpha$  flux profile, as in Prochaska et al. (2013). Indeed, averaging their estimate of the cross-correlation over the the line-of-sight bins corresponding to a velocity window of  $\pm 1000 \text{ km s}^{-1}$ , we can infer the corresponding mean flux fluctuations as a function of the transverse separation between QSOs. We plot the resulting  $\delta_F$  profile in Figure 3.4 with orange triangles, and list them in Table D.2 in the Appendix D, where we also provide the details of our conversion. The vertical bars are the  $1\sigma$  errors of the measurements, whereas the horizontal bars show the bin widths. For the first time, we show that Font-Ribera et al. (2013) measurements are consistent with Prochaska et al. (2013), extending the dynamic range probed by the Ly $\alpha$  absorption lines from the CGM out to the IGM. As such, these measurements have the potential to jointly constrain the physics of both CGM and IGM.

We plot the results of the simulations in Figure 3.4, with a point in the center of each impact parameter bin. To guide the eye, we linearly interpolate between the points. The magenta diamonds and the green circles, connected with solid lines of the same

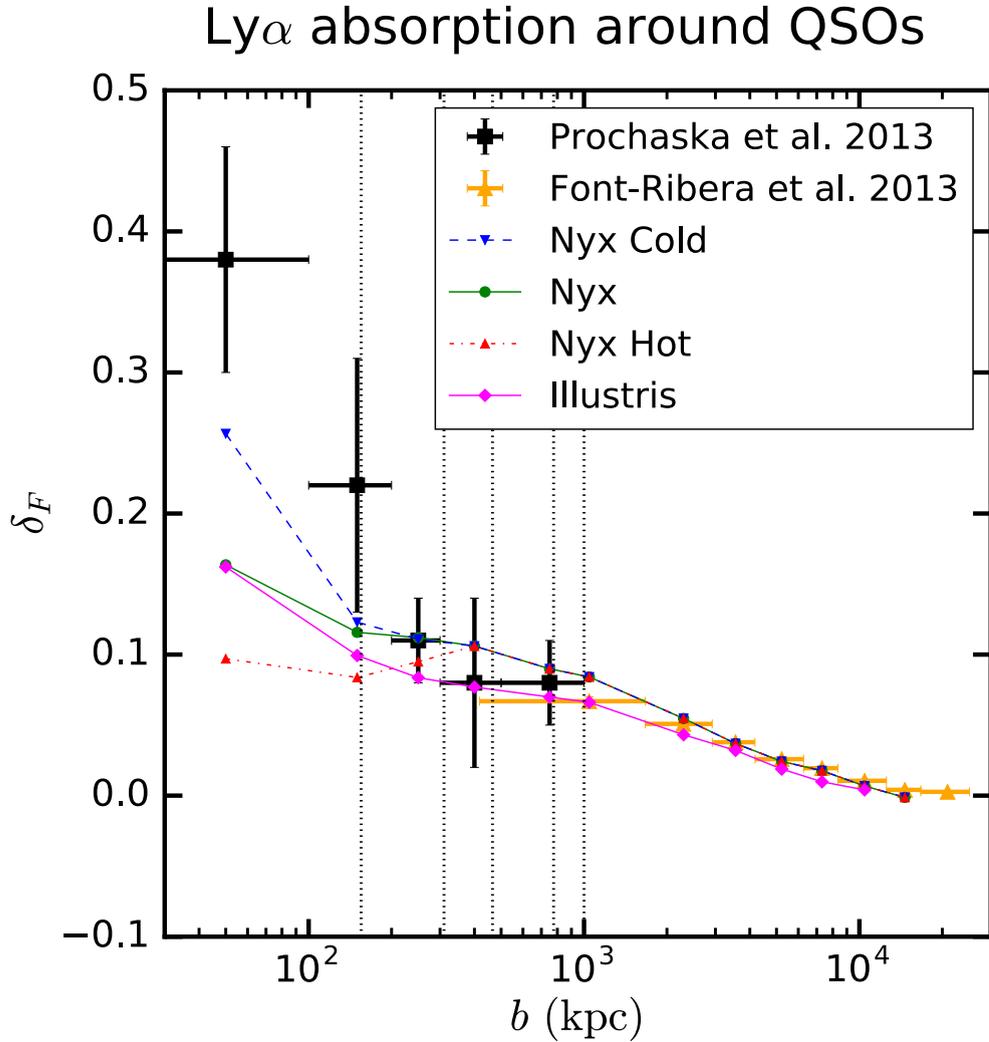


FIGURE 3.4: Mean Ly $\alpha$  flux fluctuations at different impact parameter bins ( $b$ ) from a foreground QSO ( halo mass  $M \gtrsim 10^{12.5} M_\odot$ ), with respect to the mean flux in the IGM. The black squares are the measurements by Prochaska et al. (2013), while the orange triangles are obtained from the measurements of the QSO-Ly $\alpha$  cross-correlation function by Font-Ribera et al. (2013). The magenta diamonds and green circles, linearly interpolated with solid line with the same colors, are the results obtained with Illustris and Nyx, respectively. The red triangles connected with the dot-dashed red line and the blue reversed triangles linearly interpolated with the blue dashed line refer to the Nyx Hot and Nyx Cold models, respectively (see text for details). The vertical black dotted lines mark 1, 2, 3 and 5 times the virial radius corresponding to the minimum mass of the sample of halos considered in Nyx ( $10^{12.5} M_\odot$ ), as well as the 1 Mpc boundary (see the discussion in § 3.4.1.3). While being consistent with the observations outside the virial radius, all simulations struggle at reproducing the data at small impact parameter, indicating a lack of HI absorption in the CGM of quasar hosts in the simulations considered. At 1 Mpc, Illustris matches the observations, while Nyx overshoots them. Thus, BOSS data are already capable of tightly constraining simulations.

colors, refer to Illustris and the pure Nyx run, respectively. The red triangles connected with a dot-dashed red line and the blue reversed triangles interpolated with a dashed blue line represent the Nyx Hot and Nyx Cold models, respectively. The simulations do not extend to the outermost data points because the size of their boxes is not large enough. The vertical dotted black lines correspond to 1, 2, 3 and 5 times the virial radius corresponding to the minimum mass of the sample of halos considered in Nyx ( $10^{12.5} M_{\odot}$ ), as well as the 1 Mpc boundary (it will be useful in the discussion in § 3.4.1.3). The mean flux fluctuations predicted by the simulations in each impact parameter bin is obtained averaging the values of  $\delta_F$  obtained from 5 different samples of  $10^4$  skewers, selected as explained in § 3.2.2. We verified that the scatter in the predictions given by the various samples of  $10^4$  skewers is  $< 7\%$ .

All models considered are consistent with the observations outside the virial radius, except that Nyx overshoots the data point at 1 Mpc, which is instead matched by Illustris. Conversely, on scales  $> 2$  Mpc, the data prefer the Nyx model. For  $b \gtrsim 400$  kpc, the Nyx Hot and Nyx Cold models give exactly the same predictions as Nyx because, by construction, the temperature-density relationship is modified only within the virial radius of the foreground object<sup>5</sup>. No Nyx-based model nor Illustris reproduce the mean flux fluctuations within the virial radius. This result underlines that the simulations considered do not produce enough HI absorption in this range of distance. One possible reason is that the simulated gas in the CGM is too hot. Indeed, the Nyx Cold model is much closer to the data than the Nyx and Nyx Hot ones. Nevertheless, this is only one possibility to explain the discrepancy. In § 3.4 we shall discuss other possible reasons.

Nyx and Illustris give different predictions for the mean flux fluctuations between 200 kpc and 2 Mpc from the foreground QSO, well beyond the median virial radius of the sample. We think that they stem from the different temperature-density relationship of the gas in the CGM of Illustris and Nyx halos. We shall discuss this point into detail in § 3.4.1. The difference between the two simulations is particularly pronounced in the range  $200 \text{ kpc} \lesssim b \lesssim 2 \text{ Mpc}$ . Whereas for  $200 \text{ kpc} \lesssim b \lesssim 500 \text{ kpc}$  the error bars in Prochaska et al. (2013) measurements are too large to rule out either simulations, the data point at 1 Mpc validates Illustris, while being discrepant with Nyx. This result shows that, with the current exquisitely precise Ly $\alpha$  absorption BOSS data, it is already possible to tightly constrain simulations. The precision of the observations at smaller impact parameters (Prochaska et al., 2013) is instead limited by relatively poor statistics of the QSO pairs. In fact, the QSO sample in Font-Ribera et al. (2013) is  $\sim 100$  times larger than the sample in Prochaska et al. (2013). Therefore, if future observations (with e.g. JWST<sup>6</sup>, DESI, 4MOST<sup>7</sup>) increased the statistics of the QSO pairs with

<sup>5</sup>The minimum mass of the sample is  $10^{12.5} M_{\odot}$ , corresponding to 140 kpc. Since the sample contains halos as massive as  $10^{13.7} M_{\odot}$ , the Nyx Hot and Nyx Cold can differ with respect to Nyx up to  $b \sim 350$  kpc. As we mentioned in Footnote 4, our semi-analytic technique can affect the temperature in the virial radius of the selected halos up to  $\sim 1.2$ , and this explains the differences observed in Figure 3.4 up to  $b \sim 400$  kpc.

<sup>6</sup>James Webb Space Telescope (Gardner et al., 2006)

<sup>7</sup>4-metre Multi-Object Spectroscopic Telescope (de Jong et al., 2016)

separations in the range  $200 \text{ kpc} \lesssim b \lesssim 1 \text{ Mpc}$ , it would be possible to obtain more precise measurements of  $\delta_F$ , thus further constraining simulations. Since feedback prescriptions are expected to have a greater impact at small impact parameter, increasing the precision of observations in at  $< 1 \text{ Mpc}$  would enable us to provide an additional test to validate feedback implementations.

### 3.3.2.2 Damped Ly $\alpha$ Absorbers

We now compare the observations of Ly $\alpha$  absorption around DLAs by Rubin et al. (2015) and Font-Ribera et al. (2012b) with the predictions given by the simulations considered in this work.

Rubin et al. (2015) considered a sample of 40 DLAs in the redshift range  $1.6 < z < 3.6$ , intervening along the line of sight of a background QSO, and passing at different impact parameters from another background QSO. They stacked the absorption spectra in four impact parameter bins, and measured the equivalent width of the Ly $\alpha$  absorption feature within a velocity window of  $\pm 500 \text{ km s}^{-1}$  around the DLA. We converted the measured equivalent width in each bin into the corresponding mean flux fluctuation. The results are plotted in Figure 3.5 as black squares. The vertical bars indicate the  $1\sigma$  errors of the measurements, while the horizontal bars are the bin widths. We reported the inferred mean flux fluctuations in Table E.1 in the Appendix E, where we also explain the details of our conversion of the Rubin et al. (2015) measurement.

Font-Ribera et al. (2012b) considered a sample of  $\sim 5 \times 10^4$  QSO spectra from the 9th Data Release of BOSS, and a subsample of  $\sim 10^4$  DLAs from the catalogue by Noterdaeme et al. (2012). They measured the cross-correlation of Ly $\alpha$  forest absorption and DLAs, in different bins of transverse and line-of-sight separation. As we did for the QSO-Ly $\alpha$  cross-correlation in § 3.3.2.1, we convert the measurements by Font-Ribera et al. (2012b) into a mean flux fluctuations profile (see Appendix D for details). The results, shown as orange triangles in Figure 3.5 and reported in Table D.1 in the Appendix D, extend the observations by Rubin et al. (2015) to 16.8 Mpc. The vertical error bars represent the  $1\sigma$  errors of the measurements, whereas the horizontal bars the widths of the impact parameter bins. The BOSS measurements are much more precise than Rubin et al. (2015) observations because of the much larger DLA and QSO samples. There seems to be a statistical fluctuation in the Rubin et al. (2015) data point at  $\sim 350 \text{ kpc}$ , which appears to be inconsistent with the otherwise smooth trend that would be inferred connecting all other data points in Figure 3.5.

We overplot the predictions given by our simulations with the same marker styles, line styles and colors as in Figure 3.4. Following the same approach as in § 3.3.2.1, we determined the statistical error in the predictions given by the simulations to be  $< 1.5\%$ . Illustris does not extend out to  $\sim 20 \text{ Mpc}$  due to its box size. Nyx, Nyx Hot, and Nyx

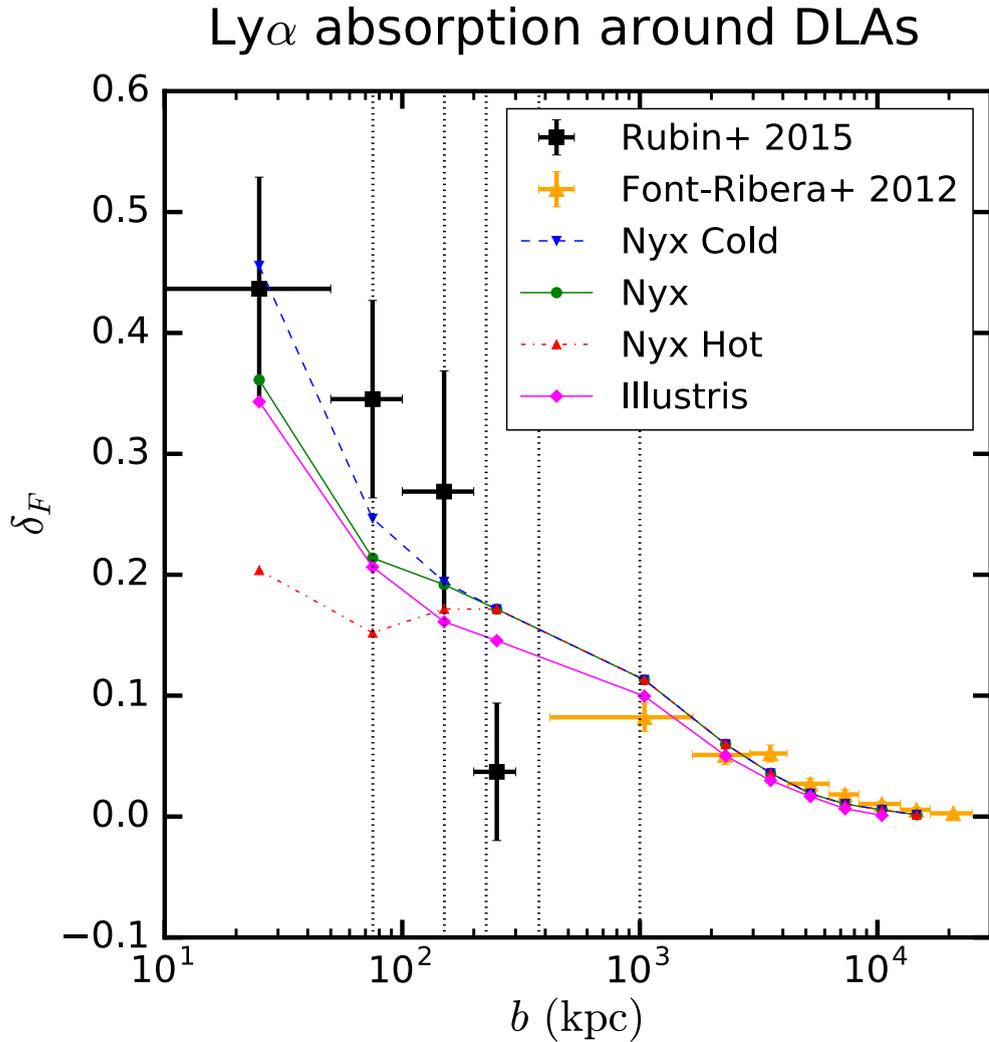


FIGURE 3.5: Mean Ly $\alpha$  flux fluctuations at different impact parameter bins ( $b$ ) around DLAs (halo mass  $M \gtrsim 10^{11.7} M_\odot$ ). The black squares are the measurements by Rubin et al. (2015), while the orange triangles are obtained from the measurements of the DLA-Ly $\alpha$  correlation function by Font-Ribera et al. (2013). The magenta diamonds and green circles, linearly interpolated with solid line with the same colors, are the results obtained with Illustris and Nyx, respectively. The red triangles connected with the dot-dashed red line and the blue reversed triangles linearly interpolated with the blue dashed line refer to the Nyx Hot and Nyx Cold, respectively (see text for details). The vertical black dotted lines mark 1, 2, 3 and 5 times the virial radius corresponding to the minimum mass of the sample of halos considered in Nyx ( $10^{11.7} M_\odot$ ), as well as the 1 Mpc boundary (see the discussion in § 3.4.1.3). Except for the Nyx Hot model, all simulations are consistent with the data. The mild tension with the data in the outermost bin of Rubin et al. (2015) can be attributed to the low number of spectra in that bin.

Cold give the same predictions for  $b \gtrsim 300$  kpc, by construction<sup>8</sup>. Except for the Nyx Hot model, all other simulations are generally consistent with the observations, both within and outside the virial radius. Apart from the already discussed data point at  $\sim 350$  kpc, there appears to be some tension with the Rubin et al. (2015) measurements at  $\sim 75$  kpc and with the Font-Ribera et al. (2012b) measurement  $\sim 1$  Mpc. Illustris is consistent with the latter within  $1\sigma$ , while Nyx within  $2\sigma$ .

We notice that the greatest differences between the predictions of Nyx and Illustris arise in the range  $100 \text{ kpc} \lesssim b \lesssim 1 \text{ Mpc}$ . Improving the precision of the measurements of the  $\delta_F$  in this range with future observations would allow setting meaningful constraints on the physics of the CGM and feedback prescriptions implemented in simulations.

### 3.3.2.3 Mean Ly $\alpha$ Transmission Profile around LBGs

We consider the measurements of the HI Ly $\alpha$  transmissivity profile in the CGM of LBGs by Adelberger et al. 2003 ( $z \sim 3$ ), Adelberger et al. 2005a (mean redshift  $z \approx 2.5$ ) and Crighton et al. 2011 ( $z \sim 3$ ). Since these measurements are at different redshifts, we need to put them on the same scale, for a fair comparison. Following Crighton et al. (2011), we re-normalize the transmissivity profile to the same mean flux  $\langle F \rangle_{\text{IGM}} = 0.76$

We convert the transmissivities into  $\delta_F$  profiles, and report them in the left panel of Figure 3.6. Grey circles, black circles and orange squares refer to the measurements by Adelberger et al. (2003), Adelberger et al. (2005a) and Crighton et al. (2011), respectively. Unlike Figures 3.4 and 3.5, the  $x$ -axis represents the 3D distance from the LBG, and not the impact parameter. In the observations, the 3D distance between each LBG and Ly $\alpha$  absorption feature is determined from their measured angular separation, and their co-moving distances from the observer. The latter are inferred from the measurement of the redshifts of the LBG and the absorption line, under the assumption of a pure Hubble flow. Therefore, the 3D distance estimated in the measurements differs from the real distance, due to velocity flows and redshift space distortions.

To reproduce the observations, we considered the halos with mass  $M > 10^{11.5} M_\odot$  in the Nyx (Illustris) snapshot at  $z = 3$  ( $z = 3.01$ ). The mass threshold has been determined as explained in § 3.2.1. Around the selected halos, we considered the same impact parameter bins adopted to reproduce the measurements by Font-Ribera et al. (2012b) and Rubin et al. (2015). We drew a sample of  $10^4$  skewers in each impact parameter bin as explained in § 3.2.2, and re-normalized the mean flux of each sample to  $\langle F \rangle_{\text{IGM}} = 0.76$  (following Crighton et al. 2011). We binned the pixels in all skewers according to their radial distance from the foreground object, using the same binning adopted in the

<sup>8</sup>The minimum mass of the sample is  $10^{11.7} M_\odot$ , corresponding to 75 kpc. Since the sample contains halos as massive as  $10^{13.7} M_\odot$ , the Nyx Hot and Nyx Cold can actually differ with respect to Nyx up to  $b \sim 350$  kpc.

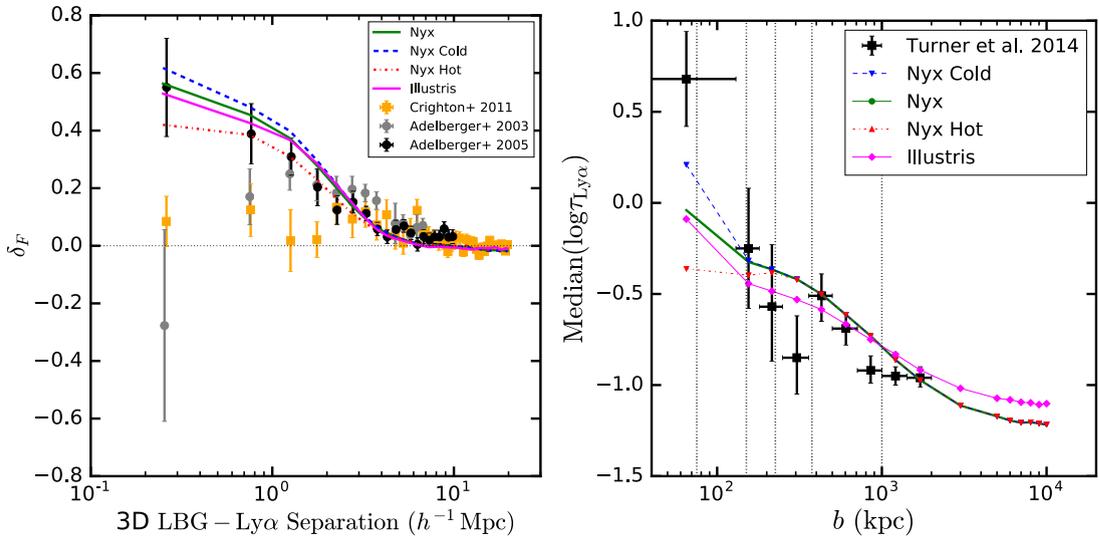
Ly $\alpha$  absorption around LBGs

FIGURE 3.6: *Left panel:* Mean Ly $\alpha$  flux fluctuations at different 3D separations from the foreground LBG (halo mass  $M \gtrsim 10^{11.7} M_{\odot}$ ). The orange squares, grey and black points are the measurements by Crighton et al. (2011), Adelberger et al. (2003) and Adelberger et al. (2005a), respectively. The solid magenta, solid green, dot-dashed red and the dashed blue lines Illustris, Nyx, Nyx Hot and Nyx Cold models, respectively (see text for details). *Right panel:* Median Ly $\alpha$  optical depth of HI in the CGM of foreground LBGs (halo mass  $M \gtrsim 10^{11.7} M_{\odot}$ ) illuminated by a background quasar. The black squares are the measurements by Turner et al. (2014). The magenta diamonds and green circles, linearly interpolated with solid lines of the same colors, are the results obtained with Illustris and Nyx, respectively. The red triangles connected with the dot-dashed red line and the blue reversed triangles linearly interpolated with the blue dashed line refer to the Nyx Hot and Nyx Cold models, respectively (see text for details). The vertical black dotted lines mark 1, 2, 3 and 5 times the virial radius corresponding to the minimum mass of the sample of halos considered in Nyx ( $10^{11.7} M_{\odot}$ ), as well as the 1 Mpc boundary (see the discussion in § 3.4.1.3).

observations that we want to reproduce. We then computed the mean flux fluctuations in each radial bin.

We plot the predictions given by the simulations and models considered in this work in the left panel of Figure 3.6. The results of Illustris, Nyx, Nyx Cold and Nyx Hot are plotted with the solid magenta, solid green, dashed blue and dot-dashed red lines, respectively. We verified that the scatter in  $\delta_F$  arising from the choice of different samples of skewers is negligible. The Nyx Cold and Nyx Hot models give different predictions than Nyx at distances  $\lesssim h^{-1} 4$  Mpc. This may seem somewhat puzzling, since in the Nyx-painted models the temperature of the CGM is altered only within the virial radius. However, the extent of the region along the flux skewers which is affected by the altered Nyx models is of the order of Mpc (see Figure 3.3). This is then consistent with the behavior of the Nyx Hot and Nyx Cold Models in the left panel of Figure 3.6.

All models are consistent with all observations at separations  $\gtrsim 5h^{-1}$  cMpc, except for the tension with Adelberger et al. (2005a) at  $\gtrsim 7h^{-1}$  cMpc. Between  $3h^{-1}$  cMpc and

$5h^{-1}$  cMpc the observations by Adelberger et al. (2003) are harder to reproduce, while within  $2h^{-1}$  cMpc, all models struggle reproducing Crighton et al. (2011). The innermost bin of the observations by Adelberger et al. (2005a) is consistent with all simulations, while the data between  $700h^{-1}$  ckpc and  $2h^{-1}$  cMpc are best reproduced by the Hot Nyx model.

Overall, the results in the left panel of Figure 3.6 may seem somewhat in contradiction with the findings discussed in § 3.3.2.1 and § 3.3.2.2, which were generally favoring a cooler CGM. However, it is hard to compare the radial profile of the Ly $\alpha$  transmissivity with the mean flux fluctuations as a function of the transverse separation. Indeed, at a fixed 3D distance  $R$ , one probes the Ly $\alpha$  absorption of HI at all impact parameters  $b < R$ . Therefore, radial profiles of  $\delta_F$  do not separate the physical effects occurring transverse and parallel to the line of sight, as it is the case for  $\delta_F$  profiles versus the impact parameter. Furthermore, a well-posed comparison among the data in Figures 3.4, 3.5 and 3.6 is not really possible unless we know the distribution of impact parameters in the observations.

A further reason why it is hard to interpret the results shown in this section is that the measurements considered are not everywhere consistent with one another within the error bars, and it is not obvious to understand which one is the most reliable. As pointed out by Crighton et al. (2011), their measurements are more precise at large separations, while the data by Adelberger et al. (2005a, 2003) should be more reliable at small distance from the LBG. The high transmissivity in the innermost bin in Adelberger et al. (2003) was interpreted as a bubble of ionized gas around the foreground LBG, but that result was contradicted by Adelberger et al. (2005a). Furthermore, Crighton et al. (2011) claimed that the error on their measurement in the innermost bin may be underestimated. In conclusion, it seems that the  $\delta_F$  profile as a function of the 3D separation from the LBG is not the optimal statistic to use in order to constrain simulations.

### 3.3.2.4 Median Ly $\alpha$ Optical Depth around LBGs

Turner et al. (2014) considered a sample of 854 foreground LBGs at redshift  $z \approx 2.4$ , and studied the Ly $\alpha$  and metal absorption in their CGM exploiting spectra of background QSOs. They determined the median Ly $\alpha$  pixel optical depth within a velocity window of  $\pm 170 \text{ km s}^{-1}$  around the LBG, as a function of the impact parameter. We report their measurements as black squares in the right panel of Figures 3.6. The vertical bars indicate the  $1\sigma$  errors in the measurements, whereas the horizontal bars the bin width.

We overplot the results of the models considered in this work at the same redshift of the observations, with the same colors, marker and line styles as in the left panel of Figure 3.6. Nyx and Illustris are generally consistent with the observations. Both simulations underpredict the optical depth in the innermost bin. The point at  $\sim 300$  kpc

falls  $1.3\sigma$  and  $2\sigma$  below Illustris and Nyx models, respectively, and there is some tension between simulations and measurements in the range (800, 1000) kpc. Once again, the observations within the virial radius favor a cooler CGM, even though our Cold Nyx model still underpredicts the data. The Nyx Hot and Nyx Cold models give the same predictions as Nyx at impact parameters  $b \gtrsim 300$  kpc, by construction.

At impact parameters larger than 2 Mpc, the simulations tend to an asymptotic value, which is the median optical depth of the IGM. At first glance, it might seem puzzling that the two simulations yield a different median optical depth for the IGM, but that does not necessarily have to be the case. In fact, even though the samples of skewers drawn from Nyx and Illustris are both tuned to yield the same mean flux, the PDF of the flux can still be different in the two simulations. We verified that this is indeed the case here.

### 3.3.3 Summary

In this section, we provided a comparison between several observations of Ly $\alpha$  absorption around different kinds of objects (QSOs, DLAs, LBGs). We converted large-scale DLA-Ly $\alpha$  and QSO-Ly $\alpha$  cross-correlation measurements (Font-Ribera et al., 2013, 2012b) into  $\delta_F$  profiles, highlighting that these data are consistent with measurements at smaller impact parameters (Prochaska et al., 2013, Rubin et al., 2015). The cross-correlation measurements have been so far fitted only with models based on linear theory (Font-Ribera et al., 2012b), so it is the first time that these observations are compared with full-box hydrodynamic cosmological simulations. Our comparison shows that the simulations agree with the data, confirming that Nyx and Illustris correctly describe the Ly $\alpha$  absorption in the IGM.

In the impact parameter range (200, 2000) kpc, Nyx and Illustris predict different mean Ly $\alpha$  flux profiles, yet they are both consistent with most of the data considered. If future observations detect more background QSO – foreground QSO/DLA pairs, the error bars of  $\delta_F$  in the range (200, 2000) kpc would be reduced, allowing us to understand which simulations provide a correct description of the Ly $\alpha$  absorption at the interface of IGM and CGM. As such, matching  $\delta_F$  data over a wide range of impact parameters has the potential of becoming a powerful test for cosmological simulations.

The mean Ly $\alpha$  flux profiles show an excess of absorption in the CGM of QSOs and DLAs, asymptotically reaching the value of mean Ly $\alpha$  flux in the IGM at  $b \approx 10$  Mpc. All simulations should reproduce such behavior, once normalized to the correct mean Ly $\alpha$  flux in the IGM. However, even if simulations are tuned to reproduce this quantity, there is no warranty that they reproduce the asymptotic median Ly $\alpha$  optical depth, as the PDF of the Ly $\alpha$  flux may be different. So, at the current stage, the median Ly $\alpha$  optical depth profile seems to be less informative than  $\delta_F$ , as far as testing cosmological simulations is concerned. Interpreting comparisons between simulations and observations of the Ly $\alpha$

transmissivity as a function of the 3D galactocentric distance is difficult, too, because such observable does not allow differentiating between physical effects occurring along the line of sight, or transverse to it.

The comparison of Nyx and Illustris with Prochaska et al. (2013) and Turner et al. (2014) shows that Nyx and Illustris do not produce enough Ly $\alpha$  absorption from HI close to the halos. The data seem to prefer our Nyx Cold model, although a cooler CGM may not be the only possible way to improve the agreement with data. We shall discuss this issue in the next section.

## 3.4 Discussion

In the previous section, we highlighted how the Ly $\alpha$  mean flux profile can potentially constrain simulations at intermediate impact parameters ( $200 \text{ kpc} \lesssim b \lesssim 2 \text{ Mpc}$ ). We also pointed out that both Nyx and Illustris underpredict the Ly $\alpha$  absorption within the virial radius of QSOs and LBGs. In this section, we connect these results with the underlying physics of the CGM and IGM, and show that the impact of possible systematic errors in our computations would not change the conclusions of this work.

In § 3.4.1 we discuss how the different predictions of the two simulations can be explained in terms of the temperature and density of the CGM. We will also show that the comparison with Ly $\alpha$  absorption data can be exploited to constrain feedback prescriptions, thus representing a fundamental test for cosmological simulations. In § 3.4.2 and § 3.4.3, we quantify the uncertainty on the predictions of  $\delta_F$  deriving from possible errors in the calibration of the halo masses in the two simulations and in the redshift of the foreground objects, respectively. In § 3.4.4 we discuss the effect of the sample size of quasar-galaxy pairs on the estimate of  $\delta_F$ . In § 3.4.5 we study the convergence of the Illustris results. Finally, in § 3.4.6 we compare the main findings presented in this Chapter with other relevant works in the literature.

### 3.4.1 Temperature and Density of the Gas

In § 3.3, we showed that painting a different temperature-density relationship on top of Nyx can significantly change the prediction of the mean flux fluctuations within the virial radius. Clearly, the temperature-density relationship in the CGM has a strong impact on the resulting Ly $\alpha$  absorption, which is worth investigating. Indeed, in the CGM the hydrogen density can overcome the self-shielding threshold ( $\sim 6.0 \times 10^{-3} \text{ cm}^{-3}$ ; see e.g. Figure 3.1), thus the corresponding Ly $\alpha$  optical depth is set not only by the  $T^{-0.7}$  dependence of the Ly $\alpha$  cross-section, but also by collisional ionization equilibrium. Since we model both mechanisms when we compute the Ly $\alpha$  absorption, investigating

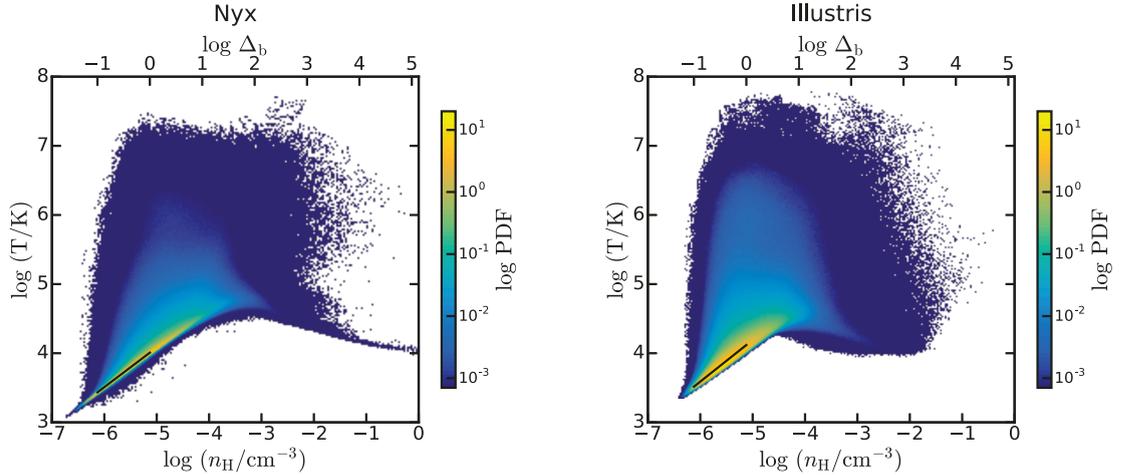


FIGURE 3.7: 2D histogram of hydrogen temperature and density for a sample of  $10^4$  skewers located at random positions in the Nyx (left panel) and Illustris (right panel) runs at  $z = 2.4$  and  $z = 2.44$ , respectively. The color bar shows the logarithm of the resulting PDF. The  $x$ -axis reports the gas overdensity (top) and the corresponding neutral hydrogen density (bottom). The black line is the best-fit power-law  $T = T_0 \Delta_b^{\gamma-1}$  across the region of the plot in the IGM regime ( $-1 < \log \Delta_b < 0$ ). The temperature-density relationships of the IGM in the two simulations are consistent with each other, meaning that they give a similar description of the IGM.

the temperature-density relationship of the gas within and outside the virial radius will provide us with good insight in the physics captured by the simulations.

### 3.4.1.1 Temperature-Density Relationship in the IGM

We want to check whether Nyx and Illustris give a consistent description of the temperature-density relationship of the IGM. We start by extracting  $10^5$  skewers in each simulation, at random positions and parallel to one side of the box. We then construct the volume-weighted 2D-histogram of density and temperature of hydrogen. In the left and right panels of Figure 3.7 we plot the results obtained for Nyx and Illustris, respectively. Although the global shape of the temperature-density relationship is similar in both simulations, Illustris presents a larger amount of hot ( $10^{5.5} \text{ K} < T < 10^7 \text{ K}$ ), rarefied ( $10^{-6} \text{ cm}^{-3} < n_{\text{H}} < 10^{-5} \text{ cm}^{-3}$ ) gas.

We divide the gas cells into two bins centered at the gas overdensities  $\Delta_{b1} = 10^{-1}$  and  $\Delta_{b2} = 1$ , with a bin width of 5% around the central value. We then compute the median temperature  $T_1^{\text{med}}$  and  $T_2^{\text{med}}$  of the gas cells in the two overdensity bins centered in  $\Delta_{b1}$  and  $\Delta_{b2}$ , respectively. Finally, we determine the power law  $T = T_0 \Delta_b^{\gamma-1}$  passing through the points  $(\Delta_{b1}, T_1^{\text{med}})$  and  $(\Delta_{b2}, T_2^{\text{med}})$ .

We obtain  $(T_0, \gamma) = (10^{4.01} \text{ K}, 1.57)$  and  $(T_0, \gamma) = (10^{4.12} \text{ K}, 1.60)$  for Nyx and Illustris, respectively. The temperature-density relationship is then very similar in the two simulations. Although Illustris and Nyx do not adopt the same model for the UVB,

the values of the photoionization rate after matching the mean flux in the IGM are very similar. At  $z = 2.4$ , it is  $1.2 \times 10^{-12} \text{ s}^{-1}$  and  $1.1 \times 10^{-12} \text{ s}^{-1}$  for Nyx and Illustris, respectively. Thus, the small differences in the temperature-density relationship arise mainly from the different reionization histories.

The fact that the temperature-density relationships are very similar in Nyx and Illustris means that the temperature-density structure of the IGM is well matched between the two simulations. The different predictions of the mean Ly $\alpha$  flux for impact parameters  $\lesssim 2 \text{ Mpc}$  suggest that this is not the case for the temperature and density of the CGM, instead.

### 3.4.1.2 Radial Temperature and Density Profiles

To begin with, we visually compare density and temperature slices around halos from both simulations. In § 3.2.4, we have already pointed out that the temperature around the Nyx halos in the first two columns from the left in Figure 3.1 broadly traces the underlying hydrogen density, except for the gas at the center of the halo, which is overcooled because of the lack of star formation.

In the third and fourth columns from the left in Figure 3.1, and in every row, we show temperature and density slices around an Illustris halo with the same mass (within 0.2%) as the Nyx halo shown in that row. The masses listed in the first two rows from the top are typical of LBG-hosting and DLA-hosting halos, whereas the halos in the remaining rows have a mass characteristic of quasar hosts. Similar to what we see in the Nyx simulation, we can clearly notice a hot region following the shape of the hydrogen overdensity. However, unlike Nyx, the size of the hot bubble tends to overcome the virial radius (indicated with a black circle around every halo) for halos with mass  $\gtrsim 10^{12.3}$  (third and fourth row of Figure 3.1. In particular, whereas the hot gas bubble in Nyx extends out to  $\sim 2 r_{\text{vir}}$  from the center the halo shown in the fourth row ( $\sim 10^{13} M_{\odot}$ ), in Illustris it reaches  $\sim 3 r_{\text{vir}}$ .

To have a deeper understanding on the extension of the hot gas phase of the CGM in Nyx and Illustris, we investigate the temperature within halos taken from the two simulations. We randomly choose 100 halos from the sample of QSO hosts given by the two simulations, which we used to reproduce the observations discussed in § 3.3.2.1. We then stack all gas cells within 1 Mpc from the centers of the halos in such subsample, and compute the median temperature within 90 equally extended bins of radial distance, normalized to the virial radius of each halo. The resulting median temperature profiles for Nyx and Illustris is shown with the black solid line in the top-left and top-middle panels of Figure 3.8, respectively. In both panels, the color bar indicates the PDF of the temperature within each radial bin. The spread of the temperature around the median is comparable in the two simulations, although it is slightly larger in Nyx ( $10^3 \text{ K} \lesssim T \lesssim 10^{7.5} \text{ K}$ ) than in Illustris ( $10^{3.4} \text{ K} \lesssim T \lesssim 10^{7.7} \text{ K}$ ). The median profile is quite

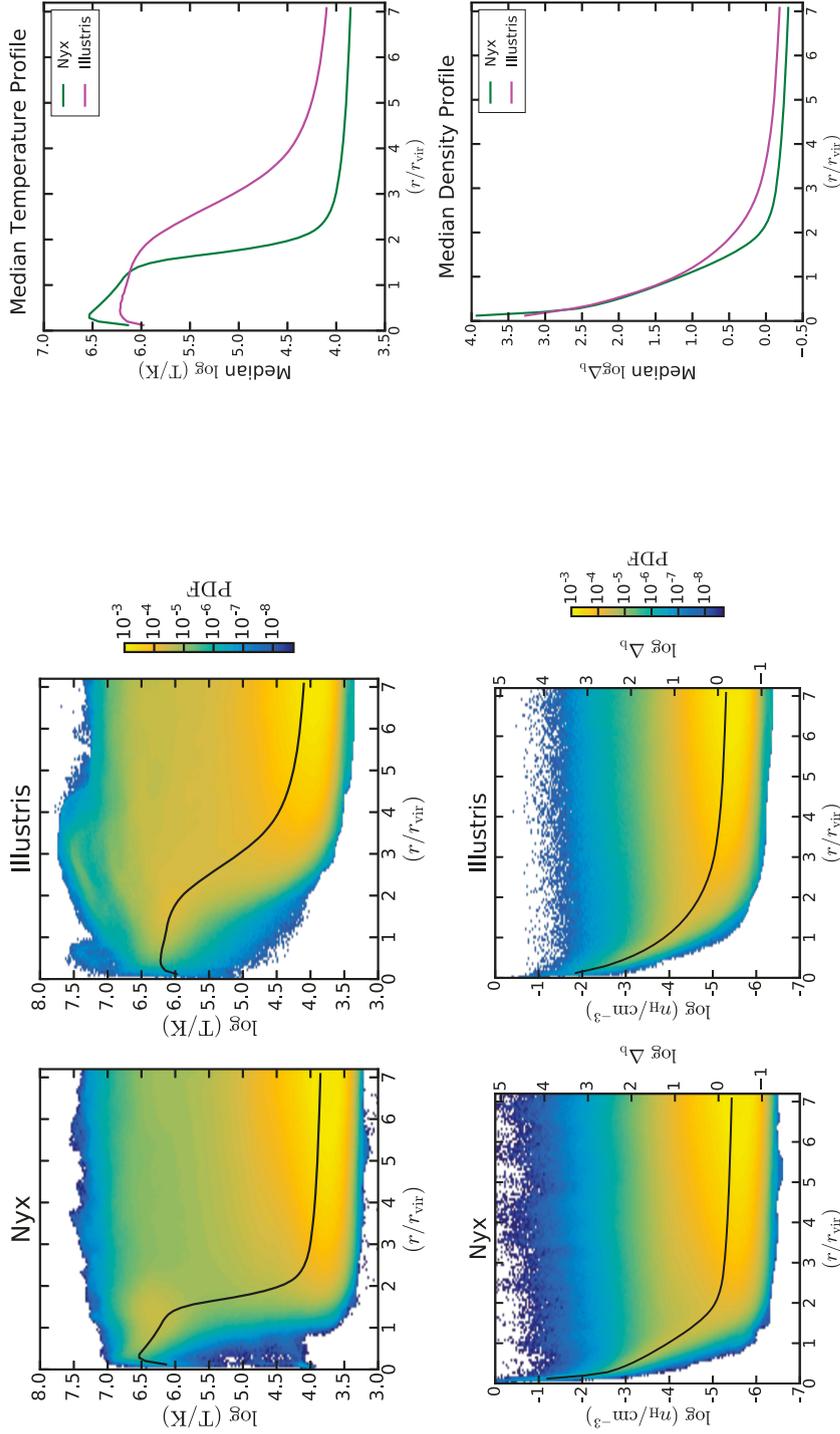


FIGURE 3.8: *Top panels:* Radial temperature profile around QSO-hosting halos in Nyx ( $M > 10^{12.5} M_{\odot}$ , left panel) and Illustris ( $M > 10^{12.4} M_{\odot}$ , middle panel). The solid black line is the median temperature profile, and the color bar gives the PDF in every radial bin within which the median is computed (see the main text for details). The median temperature profiles are plotted together in the right panel. The green and magenta solid lines refer to Nyx and Illustris, respectively. Illustris forms halos that are cooler than Nyx within the virial radius, but hotter than Nyx at larger distances. *Bottom panels:* Same as in the top panels, but for the baryon overdensity profile. Within the virial radius, Illustris is less dense than Nyx, but it is denser at larger distances.

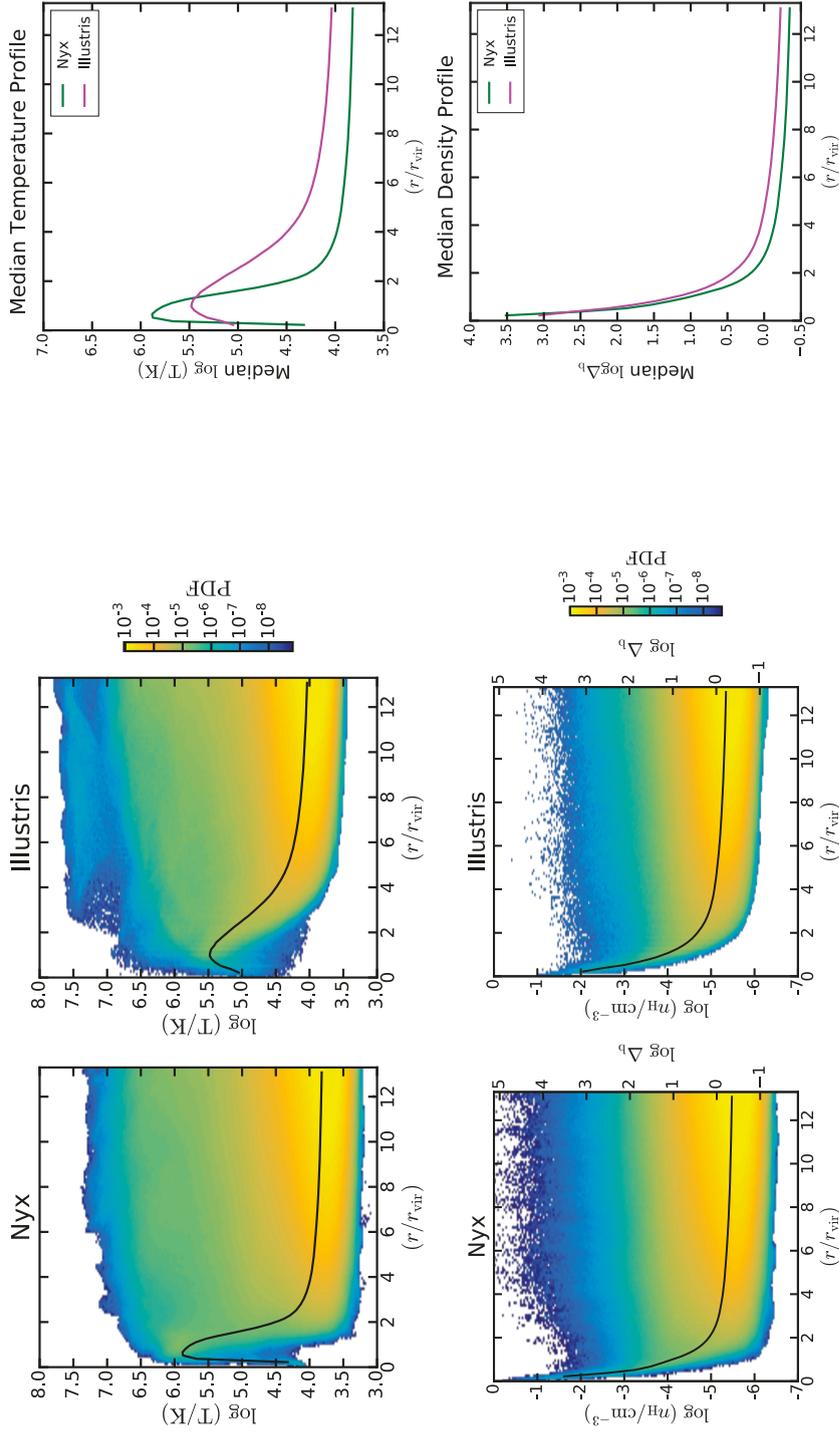


FIGURE 3.9: Same as Figure 3.8, but for LBG-hosting halos in Nyx ( $M > 10^{11.7} M_{\odot}$ ) and Illustris ( $M > 10^{11.6} M_{\odot}$ ). Also in this case, Illustris forms halos that are cooler than Nyx within the virial radius, but hotter than Nyx at larger distances.

different in the two simulations, instead. For an easier comparison, we plot it in the top-right panel of Figure 3.8. The median temperature can be up to 0.3 dex higher in Nyx (solid green line) than in Illustris (solid magenta line) within the virial radius, but in the outer region Illustris generates systematically hotter gas. The difference with respect to Nyx reaches 1.5 dex in the range  $[2r_{\text{vir}}, 3r_{\text{vir}}]$ , and is still as large as 0.3 dex out to  $7r_{\text{vir}}$ . One possible way to explain such high temperatures is the strong radio-mode AGN feedback prescription implemented in Illustris, which acts as an extra source of heating well outside the virial radius. Alternatively, the higher temperature might trace a larger underlying hydrogen density (provided it is in the regime described by the IGM power-law temperature-density relationship). It is then important to look into the density profiles of Nyx and Illustris halos, too.

The differences between the hydrogen density profiles given by Nyx and Illustris are not as pronounced as in the case of the temperature profiles. This is shown by the lower left, middle, and right panels of Figure 3.8, which display the baryon overdensity radial profiles in the two simulations, with the same structure and color coding as the corresponding top panels. Within the virial radius, Nyx and Illustris exhibit similar overdensity profiles, except for the peak within  $\sim 0.2r_{\text{vir}}$ , which is  $\sim 0.7$  dex larger in Nyx than in Illustris. This is caused by the absence of star formation in Nyx, that allows gas density to increase without converting it into stars. Outside the virial radius, Illustris is systematically denser than Nyx; the largest differences (0.4 dex) occurs at  $\sim 2r_{\text{vir}}$ , then the gap reduces down to 0.15 dex at  $7r_{\text{vir}}$ . In both simulations, the gas at  $r > 2r_{\text{vir}}$  exhibits a density typical of the IGM ( $n_{\text{H}} \lesssim 10^{-5}$ ). According to the IGM temperature-density relationship of the two simulations, a difference of 0.15 dex in  $\Delta_{\text{b}}$  should correspond to an increase of  $\sim 0.1$  dex in temperature. At face value, this could not explain the different temperature in Nyx and Illustris even as far as  $7r_{\text{vir}}$ . However, the lower left and middle panels show that there is significant scatter around the median density profile, with the bulk of gas cells lying within  $\pm \sim 1$  dex from the median value. Considering such scatter, differences in the median temperature profile of the two simulations as large as  $\sim 0.6$  dex could be explained solely in terms of their different underlying hydrogen density. On the contrary, the temperature differences as large as 1.5 dex within  $\sim 3r_{\text{vir}}$  cannot be justified with this line of reasoning, implying an extra source of heating. This strengthens our argument that the AGN feedback can be the reason behind the hotter and more extended temperature profile. From the temperature and density profiles that we obtained, we can conservatively argue that the effects of feedback in Illustris can extend out to 3 – 4 virial radii.

We repeat the same analysis shown in Figure 3.8 also for the LBG-hosting halos extracted from Nyx and Illustris. We show the results with analogous plots in Figure 3.9, which has the same structure and color coding as Figure 3.8. The behavior of the two simulations is qualitatively the same as in the case of the QSO-hosting halos. Both Nyx and Illustris exhibit colder and less dense radial profiles with respect to what we found for the QSOs, but the trend of the profiles is qualitatively similar to the case of the QSO. The largest

difference in temperature ( $\sim 0.8$  dex) occurs again at  $\sim 2 r_{\text{vir}}$ , where the hydrogen density differs by  $\sim 0.3$  dex. Although such differences are not as large as in the case of QSOs, they still underline the effect of the feedback mechanisms that are implemented in Illustris, but not in Nyx.

We now compare the temperature profiles with the mean Ly $\alpha$  flux fluctuations profiles shown in § 3.3.2.1 and § 3.3.2.2. Within the virial radius, Nyx is hotter and denser than Illustris. A higher temperature would yield more transmission, whereas a higher density would increase the optical depth. The fact that the two simulations give similar predictions for  $\delta_F$  implies that these two effects average out. Outside the virial radius, Nyx always predicts a larger Ly $\alpha$  absorption than Illustris. In this regime, most of the gas falls below the self-shielding threshold, thus photoionization is the dominant ionization mechanism, and consequently  $n_{\text{HI}} \propto T^{-0.7} n_{\text{H}}$ . Since outside the virial radius Illustris is denser than Nyx, the lower  $n_{\text{HI}}$  necessary to predict a smaller Ly $\alpha$  optical depth is set by the temperature, which is much higher than in Nyx.

From the discussion in the present subsection, we can conclude that the main driver of the differences between the  $\delta_F$  predicted by Nyx and Illustris is the temperature of the gas. In general, we can say that the mean Ly $\alpha$  flux fluctuations as a function of the impact parameter represent an excellent probe of the physics of the CGM, being closely related to the density and temperature profile of the foreground halos. As such, the  $\delta_F$  profile can be used as a further test for feedback implementations in simulations.

### 3.4.1.3 Temperature-Density Relationship in the CGM

To interpret the differences in the temperature and density profiles of halos in Nyx and Illustris, shown in § 3.4.1.2, in terms of the physics implemented in the two simulations, we repeatedly used arguments based on the temperature-density relationship of the IGM. In this section, we want to complete our discussion by investigating such relationship within different spherical shells around the center of the halos selected from the two simulations. Specifically, we consider the intervals  $(0, r_{\text{vir}})$ ,  $(r_{\text{vir}}, 2r_{\text{vir}})$ ,  $(2r_{\text{vir}}, 3r_{\text{vir}})$ ,  $(3r_{\text{vir}}, 5r_{\text{vir}})$ , and  $(5r_{\text{vir}}, 1 \text{ Mpc})$ , corresponding to the regions delimited with the vertical dotted black lines in Figures 3.4, 3.5 and in the right panel of Figure 3.6. We shall then compare the temperature-density relationships in such radial bins with the Ly $\alpha$  flux fluctuations predicted in the same intervals.

The temperature-density relationship of the hydrogen within one virial radius from the center of the QSO-hosting halos obtained in Nyx and Illustris ( $> 10^{12.5} M_{\odot}$  and  $> 10^{12.4} M_{\odot}$ , respectively) can be seen in the top-leftmost and bottom-leftmost panels of Figure 3.10, respectively. We plot the volume-weighted 2D histograms resulting from stacking 100 halos, randomly chosen from the halos used to reproduce the observations of Ly $\alpha$  absorption around QSOs (Font-Ribera et al., 2013, Prochaska et al., 2013). For both simulations, the temperature-density relationship is profoundly different from the

one of the IGM (Figure 3.7), as expected. Collisional ionization, self-shielding, and all physical processes occurring in the CGM dominate within the virial radius. Their signature in the  $T - n_{\text{H}}$  diagram is the cloud at high temperature spanning the density range  $(10^{-5}, 10^{-1}) \text{ cm}^{-3}$ , appearing in both panels. Such densities are larger than the ones present in the IGM ( $n_{\text{H}} \lesssim 10^{-5} \text{ cm}^{-3}$ ). Nyx exhibits a high-density ( $> 10^{-3} \text{ cm}^{-3}$ ), low-temperature ( $< 10^{-4.5}$ ) line, arising from the aforementioned overcooling due to the absence of star formation. Such a feature is of course absent in Illustris, which does include star formation. In addition, a minimal fraction of the gas within the virial radius in Nyx follows a low-density ( $< 10^{-4} \text{ cm}^{-3}$ ), low-temperature ( $< 10^{-4.5}$ ) line, which cannot be found in Illustris. This gas has not been shock heated, thus it lies along the temperature-density power law of the IGM. Nevertheless, the temperature-density relationship within the virial radius is qualitatively similar in the two simulations. Likewise, they predict a similar  $\delta_F$  in the same region (see Figure 3.4).

In the second panels from the left in Figure 3.10, we show the temperature-density relationship in the interval  $(r_{\text{vir}}, 2r_{\text{vir}})$ . The results from Nyx and Illustris are reported in top and bottom panels, respectively. There is a huge difference between the two simulations. Many Nyx pixels fall on the best-fit power law of the IGM (see also the left panel of Figure 3.7), but there is still a considerable fraction of gas cells in the hot phase. Illustris does not present any power-law feature and its temperature-density relationship is still dominated by the cloud of hot gas observed within one virial radius. As previously argued, the AGN feedback prescription in Illustris is probably responsible for heating the gas as far as two virial radii, erasing the IGM power-law feature in the temperature-density relationship in the range  $(r_{\text{vir}}, 2r_{\text{vir}})$ . We notice that, in the same range, the predictions of  $\delta_F$  given by Nyx and Illustris differ by  $\sim 30\%$  (see Figure 3.4). Illustris predicts less absorption, due to the larger amount of hot gas.

We plot the temperature-density relationship in the range  $(2r_{\text{vir}}, 3r_{\text{vir}})$  in the third panels from the left in Figure 3.10. Nyx (top panel) shows the typical power-law feature of the IGM. Such feature begins to appear also in Illustris (bottom panel), but the majority of the pixels still lie in the hot phase. This indicates that the AGN feedback prescription in Illustris dominates the temperature-density relationship even in the range  $(2r_{\text{vir}}, 3r_{\text{vir}})$ . The large amount of gas leads to predicting a lower absorption if compared with Nyx (see Figure 3.4).

The fourth and fifth panels from the left in Figure 3.10 show the temperature-density relationship in the radial bins  $(3r_{\text{vir}}, 5r_{\text{vir}})$  and  $(5r_{\text{vir}}, 1 \text{ Mpc})$ , respectively. In the former, Illustris still exhibits a larger amount of hotter gas than Nyx. In the latter, the diagrams of the two simulations look similar. As a reference, in the rightmost panels of Figure 3.10 we plot the temperature-density relationship of the IGM, already shown in Figure 3.7. While Nyx starts qualitatively resembling the shape of the IGM temperature-density relationship for  $r > 2r_{\text{vir}}$ , Illustris does so for  $r > 3r_{\text{vir}}$ . This is consistent with our estimate of the distance from the center of the halos up to which the

AGN feedback prescription in Illustris seems to dominate the thermal state of the CGM (see § 3.4.1.2). Therefore, also the temperature-density diagrams within different radial shells, presented in the current section, are a potentially excellent method to visualize the impact of feedback implementations on the physics of the CGM.

We repeat the same analysis discussed above for the halos hosting LBGs (and DLAs) in both simulations. The corresponding temperature-density relationships are reported in Figure 3.11. The panels report the volume-weighted 2D histograms resulting from stacking 100 halos, randomly drawn from the LBG-hosting ( $> 10^{11.7} M_{\odot}$  and  $> 10^{11.6} M_{\odot}$  for Nyx and Illustris, respectively) used to reproduce the observations of Ly $\alpha$  absorption around DLAs and LBGs (Adelberger et al., 2005a, 2003, Crighton et al., 2011, Font-Ribera et al., 2012b, Rubin et al., 2015, Turner et al., 2014). Qualitatively, the diagrams present the same differences observed for the QSO-hosting halos (Figure 3.10). Likewise, the predictions of  $\delta_F$  in the innermost bins around DLAs differ by  $\lesssim 10\%$  (see Figure 3.5). The difference increases up to  $\sim 30\%$  in the intervals  $(r_{\text{vir}}, 2r_{\text{vir}})$  and  $(2r_{\text{vir}}, 3r_{\text{vir}})$ , decreasing at farther distances.

In this subsection, we showed that, both for QSOs and DLAs, the discrepancies between the predictions of the two simulations visible in Figure 3.5 reflect the differences between the corresponding temperature-density relationships. Moreover, the temperature-density relationship within different radial shells around both kinds of objects provides valuable information about the physics of the CGM implemented in simulations.

#### 3.4.1.4 Summary

To summarize the main points of the discussion in the present section, the  $\delta_F$  profiles predicted by the simulations can be easily interpreted in terms of their underlying temperature-density relationships in the CGM. Therefore, plotting such relationship in different shells around halos is a good diagnostic of the effect of feedback prescriptions. The radial temperature and density profiles around halos in different mass ranges is an equally powerful tool to understand the physics of the CGM and relate it to the Ly $\alpha$  absorption profile, as well as to assess the impact of feedback prescriptions in simulations. The success of this approach in getting insight into the physics of the CGM and IGM stems from the simplicity of the physics behind Ly $\alpha$  absorption. Unlike the optical depth profiles of metals (e.g. CIV, OVI etc.), the Ly $\alpha$  optical depth only depends on the cosmological model, the abundance of hydrogen, and ionization equilibrium. The dependence of the recombination and ionization rates on hydrogen density and temperature are well understood, so that the  $\delta_F$  profile can be used as an additional powerful constraint for simulations. Whereas feedback implementations are generally tuned to reproduce observations directly linked with the star formation history (e.g., the observed star formation efficiency, Behroozi et al. 2013, Guo et al. 2011, Moster et al. 2013), at the same time they should be able to match Ly $\alpha$  absorption in the CGM, too.

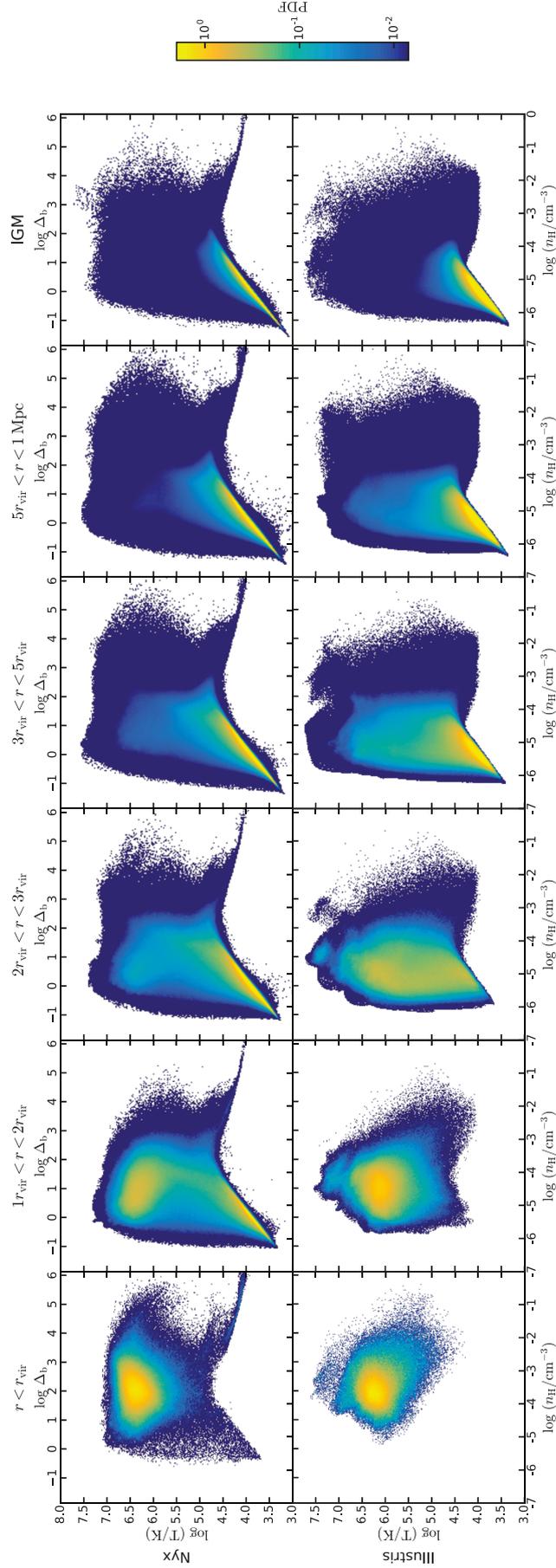


FIGURE 3.10: Temperature - density relationship of hydrogen at different radial bins from a foreground QSO. The volume-weighted 2D histograms are plotted upon stacking 100 halos, randomly chosen among the QSO-hosting Nyx halos ( $> 10^{12.5} M_{\odot}$ , top panels) and Illustris halos ( $> 10^{12.4} M_{\odot}$ , bottom panels), in different bins of distance from the center of the halos. From left to right, the histograms refer to the radial bins ( $0, r_{\text{vir}}$ ), ( $r_{\text{vir}}, 2r_{\text{vir}}$ ), ( $2r_{\text{vir}}, 3r_{\text{vir}}$ ), ( $3r_{\text{vir}}, 5r_{\text{vir}}$ ), and ( $5r_{\text{vir}}, 1 \text{ Mpc}$ ). As a reference, the temperature-density relationship in the IGM (already shown in Figure 3.7) is plotted in the rightmost panel for both simulations. The two simulations yield similar temperature-density diagrams within the virial radius, while they strongly differ in the other bins. Whereas Nyx converges to the temperature-density relationship of the IGM already in the bin ( $r_{\text{vir}}, 2r_{\text{vir}}$ ), Illustris begins producing the IGM feature only for  $r > 3r_{\text{vir}}$ .

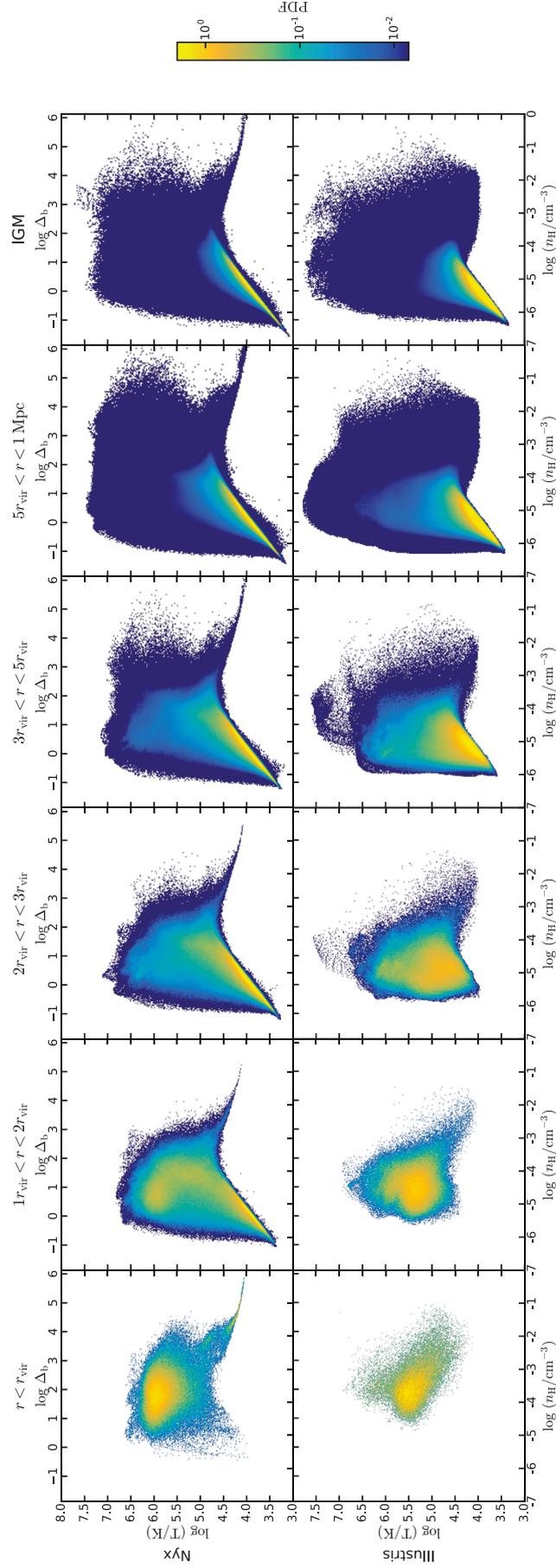


FIGURE 3.11: Same as in Figure 3.10, but for DLA/LBG-hosting halos ( $> 10^{11.7} M_{\odot}$  in Nyx and  $> 10^{11.6} M_{\odot}$  in Illustris). Whereas Nyx converges to the temperature-density relationship of the IGM already in the bin ( $r_{\text{vir}}, 2r_{\text{vir}}$ ), Illustris begins producing the IGM feature only for  $r > 3r_{\text{vir}}$ .

The exquisite precision of BOSS data favors Illustris over Nyx at 1 Mpc, highlighting the constraining power of the Ly $\alpha$  absorption around QSOs and DLAs. In the range (200 kpc, 500 kpc), both simulations are consistent with the observations within the error bars. It would be important to increase the precision of the measurements in this range with future observations, since that would allow constraining feedback prescriptions and getting more insight into the physics of the CGM.

Based on our findings, we argue that an AGN feedback implementation weaker than the one implemented in Illustris should generate fewer hot gas, making the CGM overall cooler within  $3 r_{\text{vir}}$ . Since the observations seem to favor our Nyx Cold model, a cooler CGM is a possible solution to mitigate the discrepancy between simulations and observations. On the contrary, the predicted  $\delta_F$  profiles around LBGs is consistent with data. Nonetheless, we still observe a discrepancy between simulations and Turner et al. 2014 (Figure 3.6).

In the current section, we have only considered the temperature-density relationship of the CGM and IGM. However, the Ly $\alpha$  optical depth depends also on the peculiar velocity of the gas. In principle, Nyx and Illustris may exhibit different gas velocity fields, and that may also have an impact on the Ly $\alpha$  absorption profile. We verified that the radial velocity-density and radial velocity-temperature relationships are very similar in the two simulations, confirming that the differences in the  $\delta_F$  profiles are driven by the temperature-density relationship. A more detailed discussion can be found in the Appendix F.

### 3.4.2 Halo Mass

We want to investigate how the mass of a halo impacts the Ly $\alpha$  absorption in the CGM. More massive halos reside in denser regions, which would yield more absorption. On the other hand, more massive halos contain hotter gas, and a higher temperature causes a larger transmissivity. It is not obvious which effect should prevail.

We reproduced once again the measurements of  $\delta_F$  around QSOs by Prochaska et al. (2013) and Font-Ribera et al. (2013), selecting a sample of halos within a different mass range. Instead of setting a minimum halo mass as explained in § 3.2.1, we considered three mass bins with an extension of 0.2 dex, centered in  $10^{12.2} M_{\odot}$ ,  $10^{12.4} M_{\odot}$  and  $10^{12.6} M_{\odot}$ , respectively. Likewise, to mock the observations of Ly $\alpha$  absorption around DLAs (Font-Ribera et al., 2012b, Rubin et al., 2015), we chose three mass bins centered in  $10^{11.4} M_{\odot}$ ,  $10^{11.6} M_{\odot}$  and  $10^{11.8} M_{\odot}$ , respectively, all with a width of 0.2 dex.

Both for the QSO and DLA measurements, we found out that  $\delta_F$  slightly increases at larger halo masses. This trend is in agreement with the results obtained by Meiksin et al. (2017) with the Sherwood suite of hydrodynamic simulations. At impact parameters  $\lesssim 1$  Mpc, the values of  $\delta_F$  predicted by Nyx or Illustris in two adjacent mass bins differ

by  $\lesssim 0.03$ . At larger impact parameters, the effect of the halo mass is negligible. The differences in the predictions of Nyx and Illustris for different halo masses at small impact parameters cannot explain the discrepancies between the simulations and the data. Therefore, a possible systematic error of 0.2 dex in our procedure to calibrate the halo masses (see § 3.2.1) would not affect the main conclusions of this work.

Font-Ribera et al. (2012b) estimated the typical mass of DLAs to be  $10^{12} M_{\odot}$  fitting the DLA-Ly $\alpha$  cross-correlation measurements with a model based on linear theory. This result was somewhat controversial, as DLA-hosting halos were thought to be less massive. For the first time, we compared Font-Ribera et al. (2012b) measurements with a fully non-linear model, using cosmological hydrodynamic simulations, and found that data at  $b > 2.5$  Mpc prefer larger masses, while the data points at 1 Mpc and 2.2 Mpc favor smaller masses. In any case, all mass bins considered in this section are consistent with Font-Ribera et al. (2012b) data.

### 3.4.3 Redshift of Foreground Objects

All observations mentioned in this work have been reproduced with Nyx and Illustris taking all foreground objects at the median redshift of the corresponding data sets, effectively neglecting their spread in redshift. Although the mean flux of the IGM evolves across the redshift range, this should not represent a big issue, since the quantity provided by the observations is not the mean flux profile, but the  $\delta_F$  profile. Despite the definition of  $\delta_F$  normalizes out the mean flux of the IGM, there may still be some residual redshift-dependence in the  $\delta_F$  profile predicted by the simulations, which needs to be evaluated. For this purpose, we reproduced Prochaska et al. (2013) taking all foreground QSOs at redshifts  $z = 2$  and  $z = 3$ , which bracket the redshift range of the observations. As expected, at  $z = 3$  we have more absorption, because the neutral fraction of hydrogen is higher at earlier redshifts. Nevertheless, in the innermost bin,  $\delta_F$  increases only by  $\sim 0.1$  for Nyx and  $\sim 0.05$  for Illustris, which is still not enough to reproduce Prochaska et al. (2013) data. Therefore, even if we unrealistically modeled all foreground QSOs at the upper bound of the redshift range of the observations, we could not explain the underprediction of Ly $\alpha$  absorption by the simulations.

We ran an analogous test for the measurements by Rubin et al. (2015), and our conclusions did not change with respect to what stated in § 3.3.2.3. The redshift distribution of foreground LBGs in Adelberger et al. (2003), Adelberger et al. (2005a), Crighton et al. (2011) and Turner et al. (2014) is so narrow with respect to the snapshots available for the simulations, that the test discussed in this section becomes superfluous.

There is another effect connected to the redshift of the foreground objects. The velocity windows considered to reproduce the observations are centered around the systemic velocity of the foreground galaxies or quasars. Following Meiksin et al. (2017), we modeled the typical observational errors in the redshifts of the foreground halos by

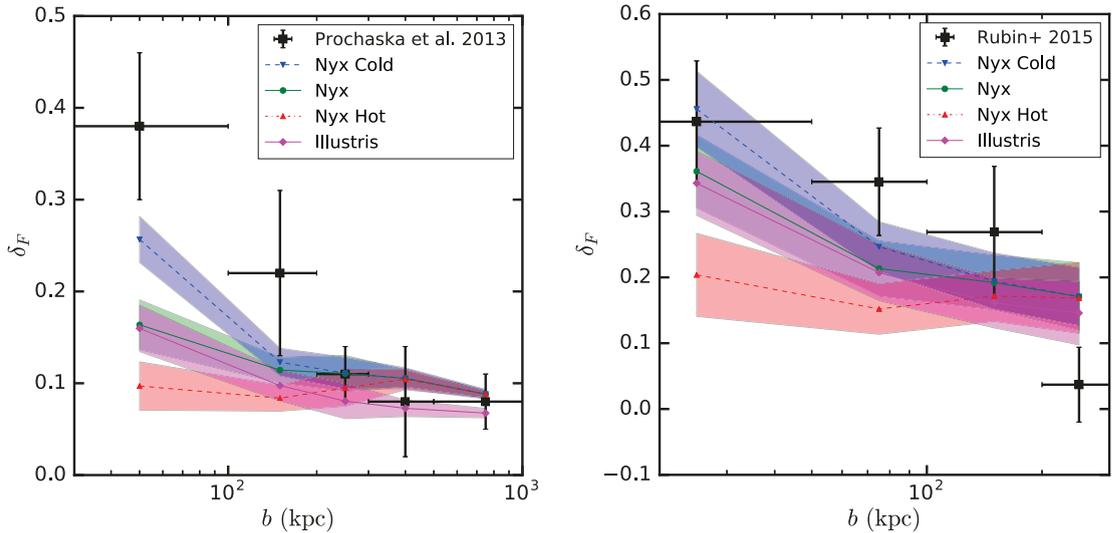


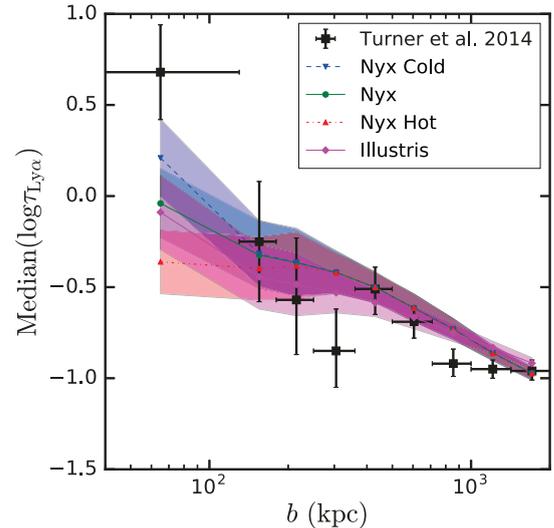
FIGURE 3.12: *Left panel:* Mean Ly $\alpha$  flux fluctuations around QSOs, at different transverse separations. The black squares represent the observations by Prochaska et al. (2013); the vertical bars are the  $1 - \sigma$  errors of the measurements, while the horizontal bars show the extension of the impact parameter bins. The results of the simulations are represented with the same color coding, markers and line styles as in Figures 3.4 and 3.5. The shaded areas delimit the  $1 - \sigma$  scatter around the estimate of  $\delta_F$  due to the limited size of the observed spectra (see text for details). *Right panel:* Same as the left panel, but for the mean Ly $\alpha$  flux fluctuations around DLAs. The black squares represent the observations by Rubin et al. (2015); the vertical bars are the  $1 - \sigma$  errors of the measurements, while the horizontal bars show the extension of the impact parameter bins.

adding a Gaussian-distributed random component to their velocities, with a variance of  $130 \text{ km s}^{-1}$  and  $520 \text{ km s}^{-1}$  for LBG/DLA and quasar hosts, respectively. We found that introducing such scatter has a marginal ( $< 1.2\%$ ) effect on the prediction of the Ly $\alpha$  absorption profiles. This is not surprising, because the data are already averaged over a large velocity window along the line of sight, exactly for the purpose of dealing with the errors on the redshifts of the foreground objects.

### 3.4.4 Sample Size of Observed Spectra

Whereas the Ly $\alpha$  absorption at large separation from foreground galaxies or quasars can be measured from tens of thousands of QSO spectra thanks to large-scale surveys like BOSS (Font-Ribera et al., 2013, 2012b), the number of the background quasar - foreground galaxy pairs with small transverse separations is about two orders of magnitude smaller (Prochaska et al., 2013, Rubin et al., 2015, Turner et al., 2014). We want to understand to what extent the poor statistics of spectra in observations can affect the error on the estimation of the mean Ly $\alpha$  flux fluctuations. This uncertainty is already accounted for by the error bars in the data, but we can use simulations to estimate its contribution to the total error in the measurements.

FIGURE 3.13: Logarithm of the median Ly $\alpha$  pixel optical depth around LBGs, at different transverse separations. The black squares represent the observations by Turner et al. (2014); the vertical bars are the  $1 - \sigma$  errors of the measurements, while the horizontal bars show the extension of the impact parameter bins. The results of the simulations are represented with the same color coding, markers and line styles as in the right panel of Figure 3.6. The shaded areas delimit the  $1 - \sigma$  scatter around the estimate of the median optical depth due to the limited size of the observed spectra (see text for details). The sample size contributes significantly to the errors of the measurements.



From our sample of simulated spectra in each impact parameter bin of the observations by Prochaska et al. (2013), Turner et al. (2014) and Rubin et al. (2015), we draw 50 subsamples with as many skewers as the observed spectra in the bin considered. We plot our results for the observations by Prochaska et al. (2013) and Rubin et al. (2015) in the left and right panels of Figure 3.12, respectively. The black squares represent the observations; the vertical bars are the  $1\sigma$  errors in the measurements, while the horizontal bars mark the bin widths. The results of the simulations follow the same color coding, marker and line styles as in Figure 3.4. The shaded magenta, green, blue and red regions delimit the 16<sup>th</sup> and 84<sup>th</sup> percentiles of the distribution of the estimate of  $\delta_F$  given by the 50 subsamples around the value obtained from the entire sample. As such, the bands represent the contribution to the error of the measured  $\delta_F$  due to the number of observed spectra.

The left panel of Figure 3.12 shows that the limited size of the sample of observed spectra contributes up to  $\sim 60 - 70\%$  to the total error of Prochaska et al. (2013) measurements in the range (200, 500) kpc, and appears to be the dominant source of error in the bin (200, 300) kpc. The contribution is even bigger in the observations by Rubin et al. 2015 (right panel of Figure 3.12) in the range (50, 300) kpc, and dominates the error bar in the innermost bin.

Figure 3.13 shows that the scatter in the median Ly $\alpha$  optical depth due to the sample size of the observations by Turner et al. (2014). The black squares represent the data, whereas the simulations follow the color-coding, lines and marker style of the lines as in Figure 3.12. Also the shaded areas have the same meaning as in Figure 3.12. The contribution to the error in the measurements due to the size of the sample of spectra is significant also in this case, accounting for  $\sim 50 - 70\%$  of the error bar.

We could not compute the scatter in  $\delta_F$  due to the sample size of spectra in BOSS (Font-Ribera et al., 2013, 2012b). We verified that, assuming that the spectra they uniformly populate the transverse separation bins, the scatter in  $\delta_F$  due to the sample size is negligible. From the size of the LBG sample in (Adelberger et al., 2003), Adelberger et al. (2005a) and Crighton et al. (2011), we expect the scatter in  $\delta_F$  to be of the same order of magnitude as in Rubin et al. (2015) (right panel of Figure 3.12).

To summarize, the simulations considered in this work predict that the relatively poor statistics of observed QSO spectra at small separation ( $\lesssim 1$  Mpc) from foreground objects should contribute up to  $\sim 50\% - 70\%$  to the errors on the measurements. Given that the error bars are dominated by the statistical error, and not by systematics (Font-Ribera et al., 2013, 2012b, Prochaska et al., 2013, Rubin et al., 2015, Turner et al., 2014), this means that the simulations considered in this work underestimate the variance of the observations.

### 3.4.5 Sub-resolution Physics

The limited resolution of simulations can be one of the reasons for the discrepancies between the simulations and observations shown in this work. Observations of Ly $\alpha$  absorption around foreground  $z \approx 2.5$  galaxies imply the presence of large-column density, metal-enriched,  $< 500$  pc clouds within an otherwise diffuse CGM (Crighton et al. 2015; see also Crighton et al. 2013, 2015, Simcoe et al. 2006). These clouds can be resolved neither by state-of-the art cosmological hydrodynamic simulations nor zoom-in numerical simulations. In fact, Crighton et al. (2015) showed that, to resolve the clumps inferred from their observations, AMR (Adaptive Mesh Refinement) simulations should reach a cell size of  $\lesssim 140$  pc in the CGM, and SPH simulations should have a mass resolution better than  $4 M_\odot$  (see also Agertz et al., 2007, McCourt et al., 2016, Stern et al., 2016). These requirements are several orders of magnitude beyond the achievements of any cosmological numerical simulation at present.

We compared the  $\delta_F$  profiles predicted by the high-resolution Illustris run adopted in this work with the profiles given by the two publicly available runs at lower resolutions. Specifically, their mass resolutions are a factor 8 and 64 worse than in the high-resolution run, respectively. We verified that the predictions of the  $\delta_F$  profile given by Illustris at the three available resolutions are well converged for  $b > 500$  kpc. The intermediate and high resolution runs agree within 5% for  $50 \text{ kpc} < b < 100 \text{ kpc}$ , and within 19% for  $b < 50 \text{ kpc}$ . Conservatively assuming that, if we could improve by another factor of 8 the resolution of the high-resolution Illustris run, the  $\delta_F$  estimate in the range  $b < 50 \text{ kpc}$  would increase by another 19%, that would still not be enough to match the data.

Although zoom-in simulations seem to capture the small-scale physics of the CGM, the resolution necessary to resolve the clumpy structure of the CGM is beyond current and near-future cosmological simulations. For such simulations, as suggested by Crighton

et al. (2015), it may be wiser to simulate the CGM implementing sub-resolution prescriptions, as it is already the case for the modeling of star formation and galactic-scale outflows.

### 3.4.6 Comparison with Previous Work

There is a large body of work considering the covering factor of optically thick absorbers, absorption profiles of metal lines, or the Ly $\alpha$  column density distribution (e.g. Bolton et al., 2017, Rahmati et al., 2015, 2016, Turner et al., 2016, on top of the relevant works already cited in the introduction of this Chapter). Nevertheless, as we primarily focus on the Ly $\alpha$  mean flux in this study, we restrict our discussion to works that considered a similar statistic.

Meiksin et al. (2017) compared the Ly $\alpha$  absorption profile around QSOs measured by Prochaska et al. (2013) with two runs of the Sherwood suite of hydrodynamic simulations. Galactic winds were implemented in only one of the two runs. The predictions of  $\delta_F$  around QSOs given by the runs with and without feedback differ up to  $\sim 170$  kpc. In our work, we find discrepancies between Nyx and Illustris out to larger impact parameters, i.e.  $\sim 2$  Mpc. The larger radius to which we see differences between the simulations considered with respect to Meiksin et al. (2017) can be due to the strong AGN feedback prescription in Illustris up to  $\sim 3 r_{\text{vir}}$  (as discussed in § 3.4.1). On top of that, we use a different criterion to select the QSO-hosting halos in the simulations. We set a select halos above a certain mass threshold ( $10^{12.5} M_{\odot}$  for Nyx and  $10^{12.4} M_{\odot}$  for Illustris) because, from their observations, Prochaska et al. (2013) could not set an upper limit to the mass of the halos hosting the QSOs of their sample, but only a lower limit. Consequently, whereas most of the halos have a mass around  $\sim 10^{12.5} M_{\odot}$ , our sample includes also halos as massive as  $10^{13.7} M_{\odot}$ . Instead, Meiksin et al. (2017) consider halos with mass between  $10^{12.2} M_{\odot}$  and  $10^{12.8} M_{\odot}$ . The median virial radius of Meiksin et al. (2017) sample of halos is then smaller than ours, therefore the signature of their feedback prescription extends out to smaller impact parameters.

The no-wind run underpredicts  $\delta_F$  for transverse separations  $\lesssim 170$  kpc. Including winds increases  $\delta_F$  by an amount which varies from  $\sim 0.01$  at  $\sim 300$  ckpc up to  $\sim 0.15$  at  $\sim 50$  ckpc. Nevertheless, the increased absorption cannot resolve the discrepancy with the data. On the contrary, the run with galactic winds matches the observation of the Ly $\alpha$  equivalent width around LBGs by Steidel et al. (2010) and Rakic et al. (2012).

Turner et al. (2017) compared the observations of the median Ly $\alpha$  optical depth due to HI around LBGs by Turner et al. (2014) with the predictions of a run of the EAGLE suite of hydrodynamic simulations. The run considered included both stochastic thermal stellar feedback and AGN feedback. To reproduce the observations, they considered samples of halos in different mass bins. The best match with the data occurs for the bin ( $10^{11.5}, 10^{12.0}$ )  $M_{\odot}$ , which is consistent with the halo mass threshold for Nyx and Illustris

derived in this work ( $10^{11.7} M_{\odot}$  and  $10^{11.6} M_{\odot}$ , respectively). Whereas the simulation yields good agreement with the observations, there is some tension with the data points at a transverse distance of  $\sim 70$  kpc,  $\sim 300$  kpc and  $\sim 800$  kpc from the foreground LBG. These data points are problematic to reproduce also for Nyx and Illustris (see Figure 3.6), so our findings are consistent with Turner et al. (2017). However, we notice that the median logarithm of the Ly $\alpha$  optical depth at  $\sim 70$  kpc obtained in Turner et al. (2017) simulations underpredicts the measurements by  $\sim 0.3$ , whereas we find that the discrepancy with Nyx and Illustris is  $\sim 0.7$ . Applying the analysis developed in this work on the EAGLE simulation may then shed light on the reasons for the better match with the data.

### 3.5 Conclusions and Perspectives

For the first time, we compared state-of-the-art hydrodynamic cosmological simulations with observations of Ly $\alpha$  absorption both in the IGM and the CGM, with the aim of investigating the physical properties of such media, as well as testing feedback models implemented in the simulations. Specifically, we focused on observations of the Ly $\alpha$  mean flux profile, and Ly $\alpha$  optical depth profile, around foreground QSOs, LBGs and DLAs passing at different impact parameters from background QSOs. The data considered span a wide range of impact parameters, from 25 kpc out to 17 Mpc, thus probing both the CGM and IGM. We ran a hydrodynamic cosmological simulation utilizing Nyx code, and used the publicly available run of the Illustris simulation, based on the *Arepo* code. The former has neither metals nor prescriptions for star formation and feedback, while the latter is endowed with metals, star formation, stellar and AGN feedback prescriptions. Hereafter, we list the main conclusions of our comparison between observations and simulations.

1. We translated the Ly $\alpha$ -QSO (DLA) cross-correlation measurements obtained from BOSS data by Font-Ribera et al. 2013 (Font-Ribera et al., 2012b) into a Ly $\alpha$  mean flux profile, showing that they extend Prochaska et al. 2013 (Font-Ribera et al., 2012b) observations at small impact parameter out to 17 Mpc – see Figure 3.4 (Figure 3.5). For the first time, we simultaneously compared all the aforementioned data with hydrodynamic cosmological simulations. Both Nyx and Illustris reproduce well the data at impact parameters  $\sim 1$  Mpc, meaning that they correctly describe the Ly $\alpha$  absorption in the IGM. Although this result may not sound surprising, it was not obvious that the simulations would have matched such exquisite data.
2. At  $\sim 1$  Mpc, Nyx overshoots the observations by Font-Ribera et al. (2013), which are instead matched by Illustris. This means that the exquisite precision of the Ly $\alpha$  absorption in BOSS data can set meaningful constraints on simulations. Between

$\sim 200$  kpc and  $\sim 2$  Mpc, well outside the virial radius of QSO- and DLA-hosting halos, Nyx and Illustris predict a different mean Ly $\alpha$  flux profile, despite being consistent with currently available data (Figures 3.4 and 3.5). The constraining power of the mean Ly $\alpha$  flux profile would then benefit from increasing the precision of data with future observations. Within the virial radius, it is challenging to reproduce Ly $\alpha$  mean flux observations of QSOs with Nyx and Illustris.

3. Inspecting the temperature and density profiles around QSO- and LBG-hosting halos in the simulations, we argue that the differences between the predictions of Nyx and Illustris out to  $3r_{\text{vir}}$  can be ascribed to the AGN feedback prescription of Illustris. Indeed, the hot component of the CGM in Illustris extends out to several virial radii from the center of the halos (Figures 3.8 and 3.9). This is the main reason why it tends to systematically predict less absorption than Nyx out to 2 Mpc. Furthermore, injecting less heat through a weaker AGN feedback prescriptions within  $3r_{\text{vir}}$  should mitigate the discrepancy between Illustris and the Ly $\alpha$  absorption data (Prochaska et al., 2013, Rubin et al., 2015, Turner et al., 2014) within the virial radius.
4. Beside the temperature and density profiles, the different predictions of the mean Ly $\alpha$  flux profiles given by Nyx and Illustris can be easily interpreted in terms of the temperature-density relationship of the gas within different radial bins from the center of the halos (Figures 3.10 and 3.11). Also these diagrams support the claim that the effects of the Illustris feedback prescriptions extend out to  $\sim 3r_{\text{vir}}$ . We point out that such temperature-density diagrams, as well as the temperature and density profiles, are a valuable tool to get insight into the feedback prescriptions implemented in simulations, which in turn impact the Ly $\alpha$  absorption profiles. Therefore, the Ly $\alpha$  flux profile provides a new fundamental observable to be matched when testing feedback prescriptions, complementary to other widely used observables, such as the star formation efficiency (Behroozi et al., 2013, Guo et al., 2011, Moster et al., 2013, e.g.).
5. We also compare the simulations with several observations of the mean Ly $\alpha$  transmissivity at different 3D radial distances from foreground LBGs (Adelberger et al., 2005a, 2003, Crighton et al., 2011). Whereas all simulations are consistent with all observations at separations  $\gtrsim 5 h^{-1}$  cMpc, there is no clear indication about which model is the most favored by the data at smaller separations (left panel of Figure 3.6). Unlike the Ly $\alpha$  mean flux or the median optical depth as a function of the impact parameter, the radial profile of the Ly $\alpha$  transmissivity does not seem to be the optimal observable to constrain the properties of the CGM, because it does not clearly separate the physics occurring along the transverse and line-of-sight directions (see § 3.3.2.3).

To understand how the discrepancy between data and simulations could be solved, and to get more insight into the role of the thermal state of the CGM on the observable HI

absorption, we also developed a simple semi-analytic technique to alter the temperature of the gas cells within the virial radius of simulated halos, which produces physically sensible mock spectra. We used it to “paint” a hotter or colder temperature within the virial radius of massive halos in Nyx. We found out that the predictions of the colder model for the Ly $\alpha$  absorption within the virial radius of QSOs and LBGs is closer to the observations (Figure 3.4 and right panel of Figure 3.6). At the same time, it does not break down the agreement between Nyx and the observations of Ly $\alpha$  absorption around DLAs (Figure 3.5).

A wrong temperature-density relationship in the CGM is not the only possible reason for the discrepancy between observations and simulations, though. For example, the presence micro-turbulence in the CGM would change the peculiar velocities of the gas and thus affect the absorption spectra, although we do not expect the velocity field to have a major impact (see the discussion in the Appendix F). Moreover, the resolution of cosmological simulations is still too low to accurately describe a multi-phase, clumpy CGM (see § 3.4.5). This sub-resolution physics may increase the absorption within the virial radius. In this work, we quantified the effect of various systematics that could affect our predictions of the Ly $\alpha$  absorption profiles, and verified that they would not change our conclusions.

We reiterate that, while current observations are matched by Illustris, but not Nyx, at  $\sim 1$  Mpc, both simulations are consistent with data at  $< 500$  kpc, while giving different predictions in this range. Thus, increasing the precision of the measurements in this range with future observations would allow setting meaningful constraints on the physics of the CGM and on feedback prescriptions of different simulations. The high-precision data of Ly $\alpha$  absorption around LBGs that will soon be provided by the ongoing COSMOS Lyman Alpha Mapping And Tomography Observations (CLAMATO) survey (Lee et al., 2014) will be useful for this purpose. In this respect, it would be interesting to repeat the analysis presented in this Chapter on the recent IllustrisTNG simulation as well (Pillepich et al., 2017). For this simulation, the AGN feedback prescription is weaker than for its predecessor Illustris, hence we would expect a better match with Prochaska et al. (2013), Rubin et al. (2015) and Turner et al. (2014) within the virial radius.

This work opens several exciting perspectives. Apart from testing different feedback prescriptions against the observations of  $\delta_F$  within  $\sim 500$  kpc, we will also run a suite of simulations with different cosmological models, while leaving the remaining physics unchanged. Indeed, the exquisite precision of BOSS at  $\sim 1$  Mpc may be able to discriminate between small variations of  $\delta_F$  given by different cosmologies. We will also investigate whether such high precision can be used to set constraints on different DM models, primordial magnetic fields, and in general any physics that is expected to have an impact on large-scale structure. We will also compute Ly $\alpha$  absorption statistics other than the mean flux, as a function of the impact parameter. For example, the PDF (see

---

also Kollmeier et al., 2003) or even the line-of-sight power spectrum may retain important information. At present, there are no measurements of this kind, but instruments like the Cosmic Origin Spectrograph (COS) and surveys such as KBSS provided us with many spectra of background QSO – foreground galaxy, which can be used to compute such statistics. We would then be able to compare them with the predictions given by simulations, and refine even further our understanding of the physics of the CGM and the IGM.

## Chapter 4

# Testing Cosmological Predictions of Kinetic Field Theory

Different cosmological models yield different predictions for the statistical properties of the distribution of matter in the Universe (e.g. Dodelson, 2003, Peebles, 1980, and references therein). Since galaxies are a tracer of this distribution, large-scale surveys of galaxy clustering provide us with a mine of data for constraining cosmology. Indeed, several galaxy surveys have been performed in the past (e.g., SuperCOSMOS surveys<sup>1</sup>, 2MASS<sup>2</sup> and 6dFGS<sup>3</sup>), some are still ongoing (e.g., SDSS<sup>4</sup>), and others are programmed in the near future (e.g., eBOSS<sup>5</sup>, DESI DESI Collaboration et al. 2016 and the Euclid mission Amendola et al. 2013, Laureijs et al. 2011). An accurate theoretical modeling of clustering is necessary to interpret the increasing amount of available data.

A systematic study of clustering was initiated by Peebles (1973), Yu & Peebles (1969) and Peebles & Hauser (1974). These pioneering works informed an entire research

---

This Chapter presents my personal contribution to an ongoing project led by Carsten Littek and supervised by Prof. Dr. Matthias Bartelmann. I am the second leading author of the project. I conducted the analysis on the Millennium Run presented in this Chapter. Carsten Littek generalized the theory by Bartelmann et al. (2016) to compute the statistics of the momentum field described in § 4.3. Prof. Dr. Matthias Bartelmann developed the theory summarized in § 4.1, which represents the conceptual foundation of this work. He also provided guidance and advice. The work presented in this Chapter is an intermediate step within the longer-term generalization of Bartelmann et al. (2016) theory to predict the statistics of the Ly $\alpha$  forest, of which I am the leading author (Sorini et al., in prep.).

Part of the text in the introduction of this Chapter contains material from my single-author publication Sorini (2017), slightly re-adapted for this thesis. To conduct the analysis presented in this Chapter, I wrote the majority of the codes by myself and I re-adapted pieces of codes originally written by Prof. Dr. Joseph F. Hennawi.

<sup>1</sup>See Hambly et al. (2004, 2001a,b,c), Parker et al. (2005)

<sup>2</sup>2 Micron All Sky Survey (Skrutskie et al., 2006)

<sup>3</sup>6dF Galaxy survey (Campbell et al., 2014, Jones et al., 2009, 2004, Springob et al., 2014)

<sup>4</sup>Sloan Digital Sky Survey (Lundgren et al. 2015; see also Eisenstein et al. 2011, Frieman et al. 2008, York et al. 2000)

<sup>5</sup>Extended Baryon Acoustic Oscillation Spectroscopic Survey (Raichoor et al., 2017)

field, focusing on understanding the two-point correlation function of matter density fluctuations (e.g. Bagla, 1998, Clerkin et al., 2015, Cole et al., 1994, Couch et al., 1993, Efstathiou et al., 1991, Fedeli et al., 2009, Hamana et al., 2001, Matarrese et al., 1997, Moscardini et al., 2002, Neuschaefer et al., 1991, Nishioka & Yamamoto, 1999, Raccanelli et al., 2014, and references therein). The two-point correlation function contains full information if the distribution of the amplitudes of density perturbations is Gaussian. This is indeed the case for the primordial density fluctuations, but galaxy clustering probes the evolved density field, which is not Gaussian. Even in this case, the two-point correlation function represents a good starting point to investigate the statistical properties of clustering.

Formally, studying the power spectrum instead of the two-point correlation function is absolutely equivalent, the former being the Fourier transform of the latter. However, this is not necessarily true when considering the estimates of these two quantities that are obtained from finite and noisy observational data. The statistical analysis of galaxy surveys through the power spectrum represents a flourishing area of research (e.g. Baumgart & Fry, 1991, da Costa et al., 1994, Einasto, 1993, Fisher, 1993, Hamilton & Tegmark, 2000, Kaiser & Peacock, 1991, Meiksin & White, 1999, Muñoz & Loeb, 2008, Nishioka & Yamamoto, 2000, Park et al., 1992, Peacock & Nicholson, 1991, Peacock & West, 1992, Strauss et al., 1990, Vogeley et al., 1992, Webster, 1977, and references therein).

The correlations in the matter density field are only part of the information that can be extracted from large-scale surveys. For example, statistical properties like the power spectrum of the peculiar velocity field (or momentum field) of galaxies and of its divergence are key quantities for the modeling of redshift space distortions (e.g. Howlett et al., 2017, Jennings et al., 2011, and references therein). Moreover, such information can be used to constrain the growth rate of density perturbations, testing General Relativity and modified gravity theories (e.g. Hwang et al., 2016, Jennings et al., 2012, Johnson et al., 2014, Li et al., 2013, and references therein). Furthermore, the primordial streaming velocities of baryons may have an impact on the BAO signature on the statistics of the Ly $\alpha$  forest and other tracers of large-scale structure; this effect could be detected in upcoming observations, for instance thanks to DESI (e.g. Hirata, 2017).

The high precision of current galaxy surveys, increasing with future missions such as Euclid (Amendola et al., 2013, Laureijs et al., 2011), demands a detailed theoretical modeling of the relevant physics. Cosmological numerical simulations represent the most accurate tool to date to predict the statistics of matter clustering (e.g., Springel et al. 2017; see Springel 2012, for a review). The main reason is that, at scales  $k \gtrsim 0.2 h \text{ cMpc}^{-1}$ , non-linear physics plays a major role in shaping the power spectrum of density fluctuations. Therefore, in that regime, analytical predictions based on the linearization of the Einstein-Boltzmann equations (governing the time evolution of photons, baryons and DM) do not represent a good approximation (Dodelson, 2003). Nevertheless, a major

interest still remains for developing analytic or semi-analytic approaches to predict statistical properties of dark matter or baryons from first principles, also in the non-linear regime. Indeed, analytic methods are generally more flexible and numerically cheaper than numerical simulations, thus exploring a wide parameter space while saving a considerable amount of computing time. Moreover, and perhaps more importantly, analytic methods contribute to improve our understanding on the role played by every relevant physical process in the transition of the Universe from its initial quasi-homogeneous phase to the highly structured configuration at present time.

There have been various attempts to derive an approximate analytic expression for the non-linear power spectrum of density fluctuations. The most straightforward way to accomplish that is perhaps perturbation theory, where the equations for the evolution of the matter density field are perturbed beyond first order (e.g., Anselmi et al. 2011, Anselmi & Pietroni 2012, Bouchet et al. 1995, Ehlers & Buchert 1997, Ma & Bertschinger 1995, Peebles 1980, Seljak 2000, Valageas 2001, Valageas et al. 2013; see Bernardeau et al. 2002 for a complete review). A different idea to address the issue is expressing the non-linear power spectrum (or correlation function) as a function of its linear counterpart, evaluated at a different scale (Hamilton et al., 1991, Jain, 1997, Padmanabhan et al., 1996, Peacock & Dodds, 1994, 1996). Another approach is assuming a power law for the shape of the primordial power spectrum and the time evolution of the scale factor, and then determine the time evolution of the scale at which perturbations become non linear. It can be shown that, as long as the density fluctuations remain linear, the power spectrum evolves in time according to a self-similar solution (Colombi et al., 1996, Davis & Peebles, 1977, Efstathiou et al., 1988, Jain & Bertschinger, 1998). A widely used semi-analytic method to describe non-linear clustering is the “Halo model” (Ma & Fry 2000, Neyman & Scott 1952, Peacock & Smith 2000, Scherrer & Bertschinger 1991, Scoccimarro et al. 2001, Seljak 2000; see Cooray & Sheth 2002 for a review). In this model, the power spectrum is split into two terms, of which one encodes the correlations of matter residing in different halos, and the other one the correlations within the same halo. The exact functional forms of such terms are tuned with fits to results of N-body cosmological simulations (Smith et al., 2003).

Recently, a novel technique, based on non-equilibrium statistical field theory, has been developed, in order to obtain a fully analytic prediction of the non-linear power spectrum (Bartelmann et al. 2016; see also Bartelmann 2015, Bartelmann et al. 2017, Dombrowski et al. 2017, Fabis et al. 2014, Kozlikin et al. 2014). The basic idea is to use the formalism of statistical physics to describe an ensemble of classical particles in the phase space. These particles would represent DM, in analogy with N-body cosmological simulations. The time evolution of the ensemble can be described analytically, as the particles are subject to Hamiltonian dynamics. The density field can then be obtained from the final configuration of the ensemble, and this allows computing the power spectrum (as well as higher-order spectra). Bartelmann et al. (2016) showed that their theory reproduces the power spectrum given by the Coyote cosmic emulator (Heitmann et al., 2009, 2014,

2010, Lawrence et al., 2010) within 10% in the range  $1 h \text{ cMpc}^{-1} \lesssim k \lesssim 3 h \text{ cMpc}^{-1}$ . This level of agreement is better than any previous analytical technique based on perturbation theory, which breaks down at  $k \sim 0.2 h \text{ cMpc}$  (Taruya et al., 2012). As such, Bartelmann et al. (2016) theory represents the first, fully analytical, accurate prediction of the power spectrum of density fluctuations into the non-linear regime (but see also the discussion in § 4.1). Moreover, considering that Bartelmann et al. (2016) approximate the gravitational interaction between particles to be small (see § 4.1 for more details), the agreement with simulations will likely be improved once this approximation is weakened (Bartelmann et al., in prep.).

Considering the promising results of Bartelmann et al. (2016) approach for the prediction of the power spectrum of density fluctuations, the time has come to extend the formalism in order to predict also other statistics capable of constraining cosmological models. One specific aim is to predict the statistics of the Ly $\alpha$  forest already encountered in Chapter 2 (Sorini et al., in prep.). Though, modeling the Ly $\alpha$  absorption in Bartelmann et al. (2016) theory requires including baryons in the formalism (Lilow et al., in prep.), as well as understanding how to properly describe the velocity (or momentum) field. In fact, as I discussed in Chapter 2, the smoothness of the velocity field can have a significant impact on the statistics of the Ly $\alpha$  forest (see also the Appendix A).

In this Chapter, I focus on the understanding of the statistical properties of the momentum field that can be obtained generalizing Bartelmann et al. (2016) theory; this is a necessary intermediate step towards the ultimate goal of predicting the statistics of the Ly $\alpha$  forest within Bartelmann et al. (2016) approach. In particular, I present my contribution to the extension of Bartelmann et al. (2016) to predict the power spectra of momentum density, of its divergence and of its and curl (Littek, Sorini et al., in prep.). I focus on computing such statistics from the Millennium Run N-body simulation (Springel et al., 2005b), which will be needed to assess the accuracy of the analytic predictions. Since the implementation of the theory into a symbolic code is still under construction, in this Chapter I will mainly test the robustness of the predictions that I obtained using the Millennium Run. All results presented hereafter should be considered as preliminary.

In § 4.1 I provide an overview of the theory by Bartelmann et al. (2016), while in § 4.2 I describe the Millennium Run. In § 4.3 I define the statistics considered in this work, and present the predictions obtained for the power spectra in § 4.4. The discussion of the preliminary results is done in § 4.5, where I also outline the next steps of this work.

## 4.1 An Overview of Kinetic Field Theory

Following Martin et al. (1973) and Forster et al. (1977), it has recently been shown that the non-equilibrium kinetic field theory (KFT) of classical particles can be reformulated resembling the path-integral approach of statistical quantum field theory (Das & Mazenko, 2012, 2013, Mazenko, 2010, 2011). Bartelmann et al. (2016) applied such formalism in the context of cosmic structure formation, giving the first fully analytic prediction of the power spectrum of matter density fluctuations, incorporating non-linear physics, down to  $k = 10 h \text{ cMpc}^{-1}$  (see also Bartelmann, 2015, Bartelmann et al., 2017, Dombrowski et al., 2017, Fabis et al., 2014, Kozlikin et al., 2014). Unlike previous analytic and semi-analytic techniques, guided by Eulerian or Lagrangian perturbation theory, Bartelmann et al. (2016) approach is not based on perturbing the Friedmann equations around a known background solution. Instead, it builds upon KFT, considering the evolution of an ensemble of particles in phase space, from which macroscopic fields (e.g., the density or velocity fields) can be derived in second instance. As such, KFT can be seen as the analytic analog of N-body numerical simulations. In this section, we give an overview of KFT, in which we describe the main concepts and logical steps, refraining from detailing the quite hefty formalism, which goes beyond the scope of this thesis (see Bartelmann et al., 2016, for details).

The starting point is the generation of a phase-space distribution of classical particles, in analogy with the initial conditions of N-body simulations. The phase-space coordinates of each particle are drawn from a probability distribution, which encodes the correlations among the positions and momenta of the particles<sup>6</sup>. Under the assumption that the initial velocity field is a gradient field and of the continuity equation, the probability distribution is fully determined by the initial power spectrum given by the cosmological model considered.

The statistical properties of the initial ensemble of particles are described by the free generating functional (or partition function)  $Z_0$ . The particles evolve following Hamiltonian dynamics, and the statistical properties of the ensemble at a certain time are encoded by the generating functional  $Z$ , which is obtained applying a multiplicative interaction operator to  $Z_0$ . Such operator is determined by the action describing the dynamics of the ensemble. Although one could in principle derive the phase-space trajectories of the individual particles from  $Z$ , the relevant quantities in the context of cosmic structure formation are macroscopic fields. In the simplest incarnation of the theory, one needs to define at least two macroscopic fields: the density and response fields. The latter is naturally defined in the formalism of the theory, and it encapsulates the effect of

---

<sup>6</sup>The expression of the probability distribution contains an exponential of the auto-correlation of particle positions and momenta, and of the cross correlation among them. To maintain the remainder of the method fully analytical, it is necessary to expand this exponential up to the second order in the momentum auto-correlations. The amplitude of these correlations is indeed  $\ll 1$  (Bartelmann et al., 2016, 2017).

the inter-particle interactions on the evolution of the ensemble. For the aforementioned fields, a corresponding operator can be defined such that, once applied to  $Z$ , it yields the desired field.

Once the macroscopic fields are obtained, one can compute the power spectrum of the density field. The power spectrum depends also on the interaction operator associated with the response field. Familiar with quantum field theory, this operator can be Taylor-expanded, giving rise to the Feynman graphs of the theory (Feynman, 1950). Bartelmann et al. (2016) computed the power spectrum expanding the interaction operator to first order, finding an agreement better than 10% with the predictions of the Coyote emulator for  $k > 1 h \text{cMpc}^{-1}$ . The most critical regime appears to be  $(0.1, 1) h \text{cMpc}^{-1}$ , where the difference with the emulator can be as large as  $\sim 30\%$  (at  $k \approx 0.5 h \text{cMpc}^{-1}$ ). Nevertheless, the performance of KFT is much better than previous methods, based on perturbation theory, which break down around  $0.2 h \text{cMpc}^{-1}$  (Taruya et al., 2012). The agreement between KFT and the Coyote emulator is expected to improve substantially including higher-order interactions, alleviating the current discrepancy at  $k \approx 0.5 h \text{cMpc}^{-1}$  (Bartelmann et al., 2016, Bartelmann et al., in prep.).

While the method discussed in this section and presented in Bartelmann et al. (2016, 2017) focuses on the power spectrum of the density field, it can be in principle extended to compute also other quantities. For example, the formalism can be generalized to obtain the momentum density field, on top of the other macroscopic fields mentioned earlier in this section. While this extension has been accomplished from the theoretical point of view, its implementation into a symbolic code is currently being carried out (Littek, Sorini et al., in prep.). Once that is finalized, KFT will be able to predict the power spectra of the momentum density fluctuations, as well as its divergence and curl. The predictions will have to be compared against numerical simulations. In the next sections, we show the power spectra computed from the Millennium Run, which we will need for this purpose.

## 4.2 Millennium Run

We briefly describe the characteristics of the Millennium Run, referring the reader to the original paper for further details (Springel et al., 2005b). The Millennium Run is an N-body cosmological simulation based on the `Gadget-2` code (Springel, 2005). Dark matter (DM) is modeled with collisionless self-gravitating particles. The point masses representing DM particles are softened with a spline kernel with a softening length of  $5 h^{-1} \text{kpc}$ , i.e. 46.3 times smaller than the mean interparticle separation. Although the simulation follows the evolution of DM particles only, their mass is corrected to take into account the contribution of baryons to the total matter density.

The gravitational force is computed through a variation of the TreePM method (Bagla, 2002, Bode et al., 2000, Xu, 1995), where short-range and long-range gravitational forces are determined through a hierarchical multipole expansion (Barnes & Hut, 1986) and with a particle-mesh (PM) method (Hockney & Eastwood, 1981), respectively. An explicit force-split in Fourier space is adopted, giving a highly isotropic force law and minimizing force errors at the force matching scale. The timestep is set with a symplectic leap-frog scheme, differentiating the potential energy into a two components, determined by the particles at small and large distances, respectively.

The initial conditions are generated by randomly sampling DM particles from a Gaussian random field, where the amplitude of the Fourier-space modes follow a Rayleigh distribution, and the phases are drawn from a uniform distribution. The underlying linear-theory power spectrum is generated with the Boltzmann code CMBFAST (Seljak & Zaldarriaga, 1996), and corresponds to a  $\Lambda$ CDM model with cosmological parameters consistent with the constraints from WMAP and SDSS (Seljak et al., 2005, Spergel et al., 2003) :  $\Omega_{\text{m}} = \Omega_{\text{DM}} + \Omega_{\text{b}} = 0.25$ ,  $\Omega_{\text{b}} = 0.045$ ,  $\Omega_{\Lambda} = 0.75$ ,  $h = 0.73$ ,  $\sigma_8 = 0.9$  and  $n_s = 1$ , with standard definitions for all quantities. Thus, the power spectrum used to generate the initial conditions incorporates also baryon features, like the BAO. The initial particle distribution is generated at  $z_{\text{ini}} = 127$ , and the particles are given an initial displacement following the Zel'dovich (1970) approximation. The volume of the simulation is  $(500h^{-1} \text{ cMpc})^3$ , and follows the evolution of  $2160^3$  particles, with a mass resolution of  $8.6 \times 10^8 h^{-1} M_{\odot}$ . In this work, we use only the snapshot at redshift  $z = 0$ .

### 4.3 Statistics Considered

In this section, we define the statistics that we compute in this Chapter. The matter density fluctuations are defined as  $\delta(\mathbf{x}) = \rho(\mathbf{x})/\bar{\rho} - 1$ , where  $\rho(\mathbf{x})$  is the matter density field and  $\bar{\rho}$  its average density. The power spectrum of matter density fluctuations  $P_{\delta}(k)$  is given by

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}') \rangle = (2\pi^3)\delta_{\text{D}}(\mathbf{k} - \mathbf{k}')P_{\delta}(k), \quad (4.1)$$

where  $\delta_{\text{D}}$  denotes the Dirac delta. The quantities considered are taken at fixed redshift, and the power spectrum depends only on the norm of the wavevector  $\mathbf{k}$  because large-scale isotropy is assumed.

We now define the momentum density fluctuation field as

$$\mathbf{\Pi}(\mathbf{x}) = \Delta(\mathbf{x})\mathbf{u}(\mathbf{x}) = \frac{\rho(\mathbf{x})}{\bar{\rho}}\mathbf{u}(\mathbf{x}), \quad (4.2)$$

where  $\mathbf{u}(\mathbf{x})$  is the field of the peculiar velocities of matter. In analogy with (4.1), the power spectrum of momentum density fluctuations  $P_{\mathbf{\Pi}}(k)$  is given by

$$\langle \mathbf{\Pi}(\mathbf{k}) \otimes \mathbf{\Pi}(\mathbf{k}') \rangle = (2\pi^3)\delta_{\text{D}}(\mathbf{k} - \mathbf{k}')P_{\mathbf{\Pi}}(k). \quad (4.3)$$

The power spectrum of momentum density fluctuations is, by definition, a tensor of rank 3, where each component  $P_{\Pi_{ij}}$  represents the cross-power spectrum of the  $i$ -th and  $j$ -th components of the momentum density fluctuations. Dimensionally, the power spectrum of momentum density fluctuations is a volume times a velocity squared or, equivalently, an energy per unit density. A more direct physical meaning is encoded by the trace of  $\mathbf{\Pi}$ , that is<sup>7</sup>

$$\text{Tr}\langle\widehat{\mathbf{\Pi}}(\mathbf{k}) \otimes \widehat{\mathbf{\Pi}}(\mathbf{k}')\rangle = \sum_i \langle\widehat{\Delta u_i}(\mathbf{k}) \widehat{\Delta u_i}(\mathbf{k}')\rangle. \quad (4.4)$$

The right hand side of the equation above is the ensemble average of the correlations of the kinetic energy density fluctuations in Fourier space. We can then define the power spectrum of kinetic energy density fluctuations  $P_{\text{KE}}(k)$  as

$$\sum_i \langle\widehat{\Delta u_i}(\mathbf{k}) \widehat{\Delta u_i}(\mathbf{k}')\rangle = (2\pi^3) \delta_{\text{D}}(\mathbf{k} - \mathbf{k}') P_{\text{KE}}(k). \quad (4.5)$$

Apart from the kinetic energy density fluctuations, one may also consider the power spectrum of the divergence and the curl of the momentum density fluctuations, which are both scalar quantities. They are defined, respectively, by the following equations:

$$\begin{aligned} \langle\mathbf{k} \cdot \mathbf{\Pi}(\mathbf{k}) \mathbf{k}' \cdot \mathbf{\Pi}(\mathbf{k}')\rangle &= (2\pi^3) \delta_{\text{D}}(\mathbf{k} - \mathbf{k}') P_{\nabla \cdot \mathbf{\Pi}}(k) \\ \langle\mathbf{k} \times \mathbf{\Pi}(\mathbf{k}) \cdot \mathbf{k}' \times \mathbf{\Pi}(\mathbf{k}')\rangle &= (2\pi^3) \delta_{\text{D}}(\mathbf{k} - \mathbf{k}') P_{\nabla \times \mathbf{\Pi}}(k) \end{aligned} \quad (4.6)$$

which can be deduced applying the properties of the Fourier transform on the divergence and the curl of  $\mathbf{\Pi}(\mathbf{x})$ .

The power spectra of kinetic energy density fluctuations, divergence and curl of the momentum density fluctuations are not independent quantities. Expressing them in their components, and exploiting the properties of the Levi-Civita tensor, it can be easily proved that

$$P_{\nabla \times \mathbf{\Pi}}(k) = k^2 P_{\text{KE}}(k) - P_{\nabla \cdot \mathbf{\Pi}}(k). \quad (4.8)$$

Despite not being independent, it is still meaningful to investigate the shape of all power spectra defined, since each one of them can be most easily compared with different measurements from large-scale surveys (e.g. Howlett et al., 2017, Jennings et al., 2011).

## 4.4 Computing Power Spectra From Simulations

We begin with computing the power spectrum of density fluctuations from the Millennium Run, to check whether we reproduce the KFT predictions with the same precision as in Bartelmann et al. (2016).

---

<sup>7</sup>Unlike the rest of the thesis, in (4.4) and (4.5) I explicitly indicated the fields in Fourier space with an overhat, for the sake of clarity.

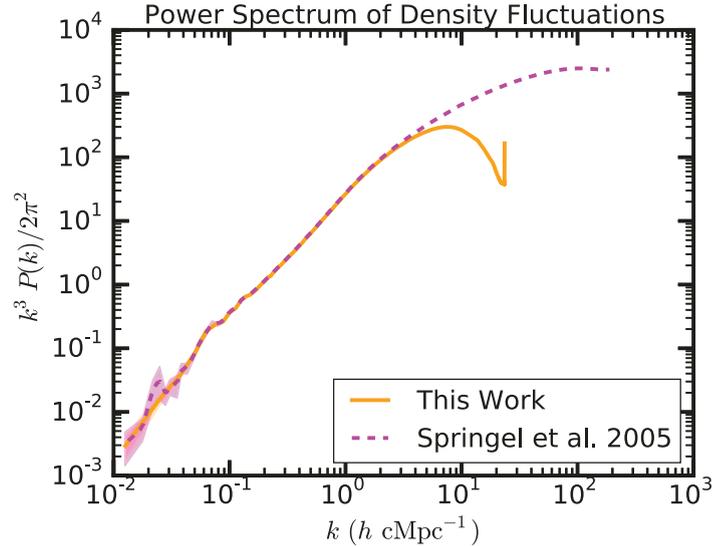


FIGURE 4.1: Dimensionless power spectrum of matter density fluctuations at redshift  $z = 0$ , obtained from the Millennium Run in this work (orange solid line) and in Springel et al. 2005b (magenta dashed line). The shaded areas around the solid orange and dashed magenta lines represent the respective scatter due to cosmic variance. Our code reproduces Springel (2005) for  $k < 2 h \text{ cMpc}^{-1}$ . The differences at larger  $k$  are due to the different algorithms implemented (see text for details).

First of all, we build the density fluctuation field from the Millennium Run by binning the dark matter particles into a regular Cartesian grid, through a CIC scheme. For each simulation, we set the size of the grid cell equal to the mean interparticle separation, i.e.  $231 h^{-1} \text{ kpc}$ . This ensures convergence of the power spectra up to  $k \sim 3 h \text{ cMpc}^{-1}$ . The detailed convergence tests are presented in the Appendix G.

To compute the power spectrum, we start by fast-Fourier transforming the density fluctuation field. We then consider a set of spherical surfaces in Fourier space, the radii of which are multiples of  $2\pi/L$ , where  $L$  is the box size.<sup>8</sup> The radii are the values of the wavenumbers  $k_s$  at which the power spectrum will be computed. The amplitude of each Fourier mode  $\delta(\mathbf{k})$  is split between the two closest Fourier-space spheres, weighted by the Fourier-space distance from such spheres. The power spectrum  $P(k_s)$  corresponding to the Fourier mode  $k_s$  is then given by the average of the square modulus of all weighted amplitudes assigned to the sphere with radius  $k_s$ .

We show the dimensionless power spectrum of density fluctuations  $k^3 P(k)/(2\pi^2)$  at  $z = 0$  that we obtain from the Millennium Run in Figure 4.1 (orange solid line). The shot noise  $P_{\text{sn}} = (L/N)^3$ , where  $N$  is the number of grid points along each side of the box, is subtracted from the power spectrum  $P(k)$ . The orange shaded area indicates the scatter due to cosmic variance. The scatter  $\Delta P(k)$  of the power spectrum at each mode

<sup>8</sup>The radius of the largest sphere in Fourier space is set to be equal to the Nyquist frequency of the grid into which the density fluctuations field is binned.

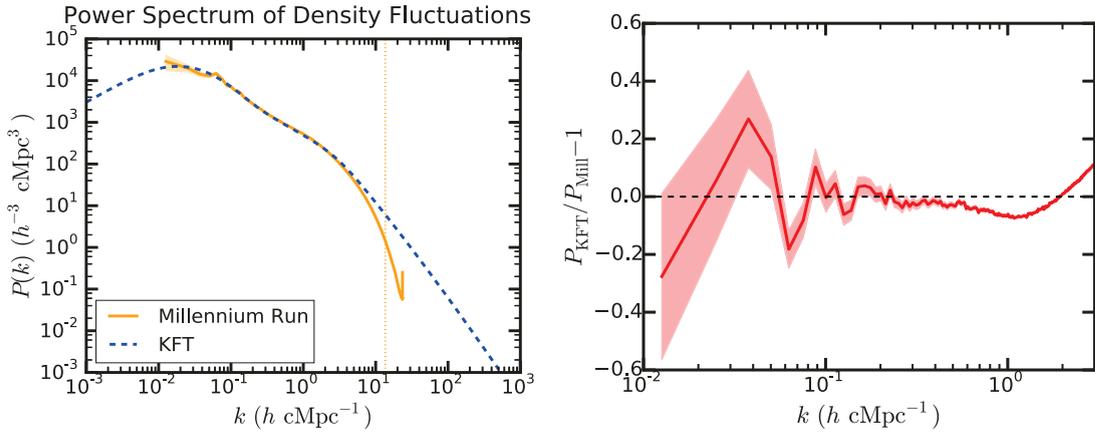


FIGURE 4.2: *Left panel:* Power spectrum of matter density fluctuations at redshift  $z = 0$ , predicted by KFT (dashed blue line) and by the Millennium Run (solid orange line). The orange shaded areas represents the error due to cosmic variance in the Millennium Run. The vertical dotted orange line is the Nyquist mode in the grid used to compute the power spectrum (see text for details). *Right panel:* relative difference between the power spectrum of matter density fluctuations predicted by KFT and by the Millennium Run. The differences are consistent with the findings in Bartelmann et al. 2016 (see text for details).

$k$  is computed as (Schneider et al., 2016)

$$\Delta P(k) = \left( \frac{2}{\Delta N_m} \right)^{\frac{1}{2}} [P(k) + P_{\text{sn}}(k)], \quad (4.9)$$

where  $\Delta N_m = L^3 k^2 \Delta k / (2\pi^2)$  is the number of modes within the bin centered in  $k$ , with width  $\Delta k$ . To check whether our code is reliable, we overplot in Fig. 4.1 the power spectrum of the Millennium Run at  $z = 0$  as published in Springel (2005), which is shown with a magenta dashed line. The magenta shaded area indicates the cosmic variance as estimated in Springel (2005).

Our code perfectly resembles Springel (2005) results for  $0.06h^{-1} \text{ cMpc} < k < 2h^{-1} \text{ cMpc}$ , and is consistent with them for  $k < 0.06h^{-1} \text{ cMpc}$ . The differences at  $k > 2h^{-1} \text{ cMpc}$  arise from the different algorithms used to compute the power spectrum. In our case we computed it on a regular Cartesian grid, treating all Fourier modes in the same way. Instead, Springel (2005) measured the power spectrum in the Millennium Run on the fly, distinguishing between “large-scale modes” and “small-scale modes”. For the large-scale modes, the power spectrum was computed through a Fourier transform of the entire box. The small-scale modes were obtained by self-folding the density field, and assuming periodicity in a fraction of the box, following Jenkins et al. (1998). The algorithm implemented in Springel (2005) is equivalent to computing the power spectrum on a regular  $81920^3$  mesh. In conclusion, our code accurately resembles the power spectrum of matter density fluctuations in the Millennium Run up to  $k \sim 2h^{-1} \text{ cMpc}$ . The differences at larger  $k$  are due to the different algorithms implemented.

Having verified that our code for computing the power spectrum is reliable, we now compare it with the power spectrum of density fluctuations predicted by KFT, choosing the same cosmological model as the Millennium Run. In the left panel of Figure 4.2 we plot the power spectrum predicted by the Millennium Run and KFT with the solid orange and dashed blue lines, respectively. The shaded orange area indicates the scatter due to cosmic variance in the Millennium Run, estimated with (4.9). In the right panel of Figure 4.2 we show the relative difference between the power spectra predicted by KFT and the Millennium Run (solid red line). The red shaded area is the scatter in the relative difference resulting from propagating the error due to cosmic variance of the Millennium Run. We investigate only modes up to  $3 h \text{ cMpc}^{-1}$ , because at larger  $k$  the convergence of our computation of the power spectrum of the Millennium Run is worse than 5%. This is due to the resolution of the grid onto which we CIC-binned the dark matter density upon computing the power spectrum (see also the Appendix G).

The predictions of KFT are overall in good agreement with the Millennium Run, in fact even better than what found by Bartelmann et al. (2016) in comparison with the Coyote emulator (Heitmann et al., 2009, 2014, 2010, Lawrence et al., 2010). In the range  $0.1 h \text{ cMpc}^{-1} < k < 3 h \text{ cMpc}^{-1}$ , the relative difference remains below 10%. In particular, for  $0.2 h \text{ cMpc}^{-1} < k < 0.5 h \text{ cMpc}^{-1}$ , well beyond the breaking point of perturbation theory (Taruya et al., 2012), the accuracy is  $\sim 3\%$ . On the other hand, for  $k < 0.1 h \text{ cMpc}^{-1}$  the difference can be as high as 30%. However, in this regime, the scatter due to cosmic variance is too large to draw robust conclusions. Indeed, the lower bound of the red shaded area in the right panel of Figure 4.2 suggests that the difference may be as small as 10%. To get more insight in this regime, it would be necessary to consider a larger simulation (e.g., the Millennium XXL Run Angulo et al., 2012), in order to reduce the uncertainty due to cosmic variance on the scales probed by the Millennium Run. This would also enable us to probe modes at larger scales ( $3 h^{-1} \text{ cGpc}$ ).

At this point, we compute the power spectrum of the momentum density fluctuations in the Millennium Run, using the same codes adopted for the power spectrum of matter density fluctuations. Specifically, in Figure 4.3 we show, from top to bottom, the power spectra of the kinetic energy density fluctuations, and of the divergence and curl of momentum density fluctuations, respectively, obtained from the Millennium Run. All power spectra are computed at redshift  $z = 0$ . In all panels, the shaded orange area is again the cosmic variance computed as in (4.9). Unfortunately, the corresponding predictions of KFT are not available yet, as the implementation of the theory into a symbolic code to generate the relevant power spectra is still in progress.

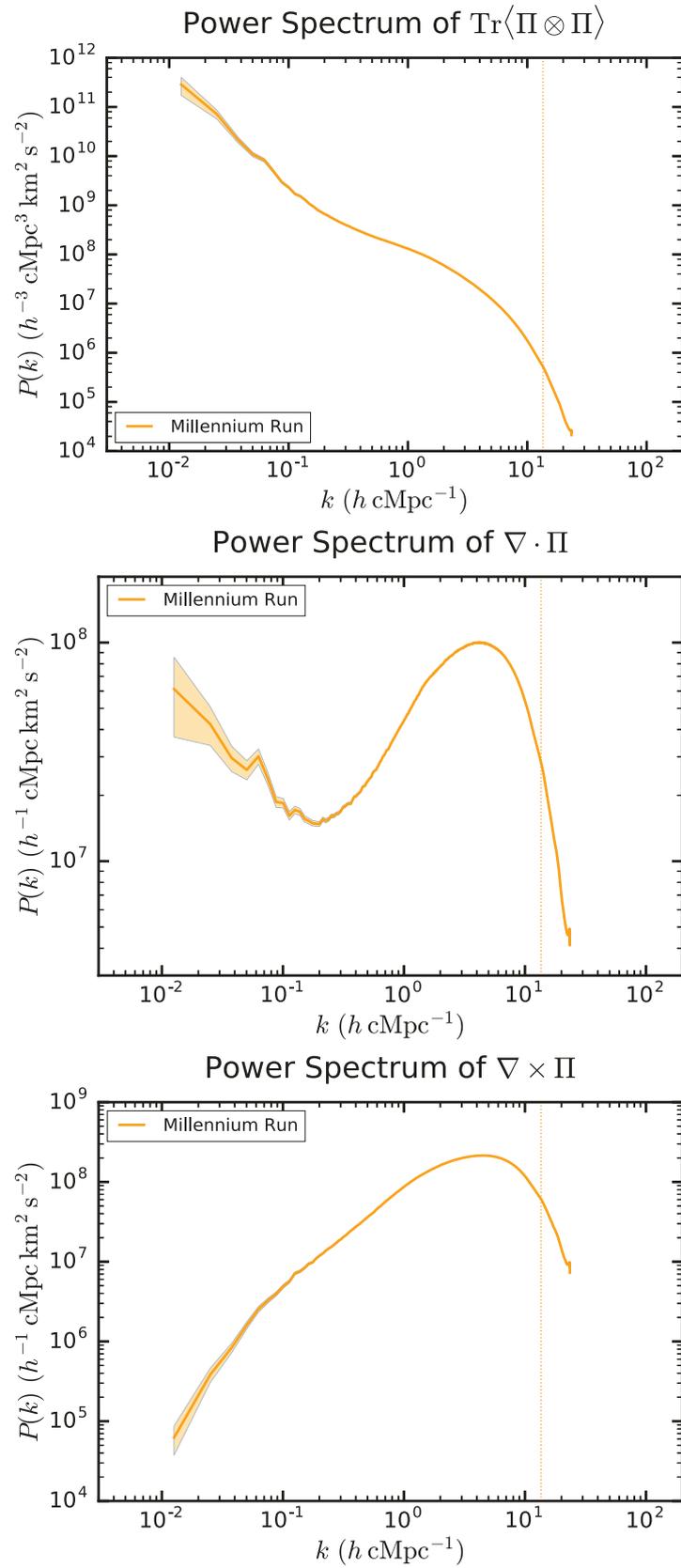


FIGURE 4.3: Power spectrum of kinetic energy density fluctuations, divergence and curl of momentum density fluctuations (top, middle, and bottom panels, respectively), predicted by the Millennium Run (solid orange line). All power spectra are computed at redshift  $z = 0$ . In all panels, the orange shaded area represents the error due to cosmic variance in the Millennium Run, while the vertical dotted orange line marks the Nyquist mode in the grid used to compute the power spectra (see text for details).

## 4.5 Discussion and Outlook

In this Chapter, we computed the power spectrum of different quantities of cosmological interest from the Millennium Run simulation, and compared it with the predictions of the KFT-based analytic approach developed by Bartelmann et al. (2016). We assessed the reliability of our code to compute the power spectrum by verifying that our results for the power spectrum of density fluctuations in the Millennium Run matches the same quantity published in Springel (2005). We also compared our estimate of the power spectrum of density fluctuations in the Millennium Run with the predictions given by KFT, with the same cosmological parameters. We showed that the agreement is better than 10% in the range  $0.1 h \text{ cMpc}^{-1} < k < 3 h \text{ cMpc}^{-1}$ , and better than 30% in the range  $0.01 h \text{ cMpc}^{-1} < k < 0.1 h \text{ cMpc}^{-1}$ , consistent with the results in Bartelmann et al. (2016).

We computed the power spectra of kinetic energy density fluctuations, as well as of the divergence and of the curl of momentum density fluctuations, in the Millennium Run. The immediate perspective of this work is finalising the implementation of a symbolic code to predict such power spectra with KFT. This will allow us to assess the performance of this theory, by comparing it with the results of the Millennium Run. The comparison would be limited to the Fourier modes sampled by the simulation ( $k > 0.0126 h \text{ cMpc}$ ). We will then consider a larger simulation, like the Millennium XXL run (Angulo et al., 2012). The box size is  $3 h^{-1} \text{ cGpc}$ , meaning that the simulation can sample the power spectrum down to  $k = 1.86 \times 10^{-3} h \text{ cMpc}^{-1}$ . Since the cosmology of the Millennium XXL is the same as the one of the Millennium Run, we will be able to use both simulations to compute the power spectra within different ranges of  $k$ , thus extending the dynamic range over which we can test KFT.

Once we verify that the predictions of the aforementioned momentum density statistics given by KFT accurately reproduce the results from N-body simulations, we can consider the treatment of the velocity field within Bartelmann et al. (2016) formalism to be reliable. Including the modeling of baryons in the theory (Lilow et al., in prep.), KFT will be ready to be generalized in order to predict the statistics of the Ly $\alpha$  forest (Sorini et al., in prep.). The predictions of the Ly $\alpha$  forest statistics will need to be compared with the results of simulations, too. For this purpose, large hydrodynamic simulations such as Nyx and Illustris (described in the previous chapters) will be considered. In order to compare the KFT predictions that we will obtain at the largest scales ( $\sim 10^{-3} h \text{ Mpc}^{-1}$ ) we would need Gpc-scale hydrodynamic simulations. As stressed in Chapter 2, there is currently no hydrodynamic simulation with such large volumes and at the same time enough high resolution to properly capture the physics of the Ly $\alpha$  forest ( $\sim 20 \text{ ckpc}$ , Lukić et al. 2015). Thus, we will apply the IMS technique developed in this thesis on top of large N-body runs, such as Millennium and Millennium XXL. Thus, this constitutes an additional perspective of the work described in Chapter 2.

---

We also point out that developing an analytic accurate method to predict the 1DPS of the Ly $\alpha$  forest will allow painting the 1D-IMS method described in Chapter 2 on large collisionless simulations, following the same logic adopted for 3D-IMS in § 2.5. This could make it possible to model the BAO signature on the Ly $\alpha$  forest in Gpc-scale N-body simulations with an even higher accuracy than the one already achieved by 3D-IMS (see § 2.7).

## Chapter 5

# Conclusions and Perspectives

The main focus of this thesis is the Ly $\alpha$  absorption line as a cosmological and astrophysical probe. At large scales, the Ly $\alpha$  forest is an exceptional observational feature that allows tracing the underlying distribution of neutral hydrogen in the IGM. As such, it can be exploited to investigate the physics of the IGM, the epoch of reionization and, more generally, constrain cosmological models (for a recent review, see McQuinn, 2016). At smaller scales, the Ly $\alpha$  absorption around galaxies at different transverse separations from background quasars allows probing the physics of the CGM and getting insight into galaxy formation (for a recent review, see Barnes et al., 2014). In this thesis, I investigated the Ly $\alpha$  absorption line both at galactic and extragalactic scales, pursuing a dual aim. In the first part of my work, I improved the modeling of the Ly $\alpha$  forest in the IGM; in the second part, I showed that, taking full advantage of observations of Ly $\alpha$  absorption both in the IGM and the CGM, it is possible to set tight constraints on state-of-the-art cosmological hydrodynamic simulations.

In chapter 2, I focused on the Ly $\alpha$  forest as a tool to probe the large-scale structure of the Universe, discussing success and limitations of cosmological hydrodynamic simulations in modeling the Ly $\alpha$  forest. Whereas such simulations led to a consistent description of the IGM in the framework of structure formation (Cen et al., 1994), they are computationally expensive, making it hard to reach both large volumes and high resolutions. On the other hand, these requirements are needed to undertake ambitious research projects, like determining the scale of the BAO from the cross-correlation of the Ly $\alpha$  forest in QSO spectra (Delubac et al. 2015, Font-Ribera et al. 2014b; see also Lukić et al. 2015). For this reason, it is of great interest to develop semi-analytic methods to predict the Ly $\alpha$  forest statistics from collisionless N-body simulations, which do not present the same computational constraints as hydrodynamic ones.

---

A minor part of the text in this Chapter is re-adapted from Sorini et al. (2016).

One of the most widely used techniques to predict the Ly $\alpha$  forest with N-body simulations consists in mocking the baryon density and velocity fields through a Gaussian smoothing of their DM analogs. The smoothing length of the Gaussian kernel  $\lambda_G$  is a free parameter, which is generally assumed to be of the same order of the Jeans scale (Gnedin & Hui 1996, Meiksin & White 2001, Rorai et al. 2013, Viel et al. 2006, 2002 but see also Peirani et al. 2014). The smoothed fields are then used to compute the Ly $\alpha$  transmitted flux within the FGPA. Before my work, there was no detailed study of the accuracy with which such method predicts the statistical properties of the Ly $\alpha$  flux, as a function of  $\lambda_G$ . This is done in Chapter 2, where we<sup>1</sup> compared the predictions of different statistics given by the Gaussian smoothing technique to the results of a reference hydrodynamic simulation. In the past literature, DM density and velocity fields were generally smoothed with the same scale  $\lambda_G$ . For the first time, we considered the possibility of having two different smoothing lengths, quantifying the impact of the smoothing scale of the velocity field on the accuracy of the method.

We showed that the accuracy of the Gaussian smoothing technique is strongly dependent on the smoothing length (see Figures 2.7 and 2.8). Since the “optimal” value of  $\lambda_G$  is guaranteed to differ for models with different IGM histories, and thus cannot be known a priori, this implies that works using *any* particular smoothing will have errors varying in an uncontrolled manner. Therefore, this thesis clarified the accuracy and the shortcomings of Gaussian smoothing, and these should be taken into account in future works making use of such technique to predict the Ly $\alpha$  forest.

To go beyond the limitations of Gaussian smoothing, we presented two new methods, 3D-IMS and 1D-IMS, based on the idea of Iteratively Matched Statistics (IMS; see § 2.3). In the former method, a mock-baryon density field is obtained from the DM density and velocity fields, such that the 3D power spectrum and PDF of the corresponding Ly $\alpha$  flux in real space, obtained within the FGPA, match by construction the same quantities given by a reference hydrodynamic simulation. The 1D-IMS method builds upon 3D-IMS, additionally matching the 1DPS and PDF of the Ly $\alpha$  flux in redshift space, given by the same reference hydrodynamic simulation. Unlike the Gaussian smoothing technique, both methods are weakly dependent on  $\lambda_G$ , meaning that they yield more robust predictions and that they can be applied to large, low-resolution N-body simulations (see Figures 2.7 and 2.8).

The 3D-IMS method is particularly suited to be applied to large N-body simulations (see § 2.5). We applied it to an  $(80 h^{-1} \text{ cMpc})^3$  N-body simulation, using a  $(10 h^{-1} \text{ cMpc})^3$  hydrodynamic simulation as calibration, and compared the results with the predictions of another hydrodynamic simulation, as large as the N-body and as resolved as the calibrating one. We showed that the average accuracy with which 3D-IMS recovers the PDF, 1DPS and 3DPS of the Ly $\alpha$  flux is better than 13%, 10% and 10%, respectively.

---

<sup>1</sup>I am referring to the first project that I undertook during my PhD, led by myself, and done together with Dr. José Oñorbe, Dr. Zarija Lukić and Prof. Dr. Joseph F. Hennawi (Sorini et al., 2016).

By comparison, the average accuracy of the Gaussian smoothing technique is 18%, 41% and 58%, respectively (see Figures 2.12 and 2.13). Therefore, in this realistic example, this thesis exposed the shortcomings of the widely used Gaussian smoothing technique, and at the same time underscored the high accuracy of the novel IMS approach presented in Chapter 2. Notably, IMS guarantees an accuracy comparable with, or better than other recent semi-analytic techniques (e.g. Peirani et al., 2014), while being simpler to implement. As such, IMS represents the state-of-the-art regarding the modeling of the Ly $\alpha$  forest in large N-body simulations.

The obvious regime of applicability of our methods stands in any project requiring large-box simulations. One example is the aforementioned modeling of the BAO signal in the Ly $\alpha$  forest. In fact, this requires simulations on a Gpc-scale box, still retaining the resolution needed to properly capture the physics of the Ly $\alpha$  forest (20 ckpc; see Lukić et al. 2015). This is well beyond the numerical capabilities of even state-of-the-art hydrodynamic simulations (e.g., Lukić et al., 2015, Pillepich et al., 2017, Schaye et al., 2015, Vogelsberger et al., 2014b). Using 3D-IMS to re-analyze the BOSS data by Delubac et al. (2015), Font-Ribera et al. (2014b) of the Ly $\alpha$  flux cross-correlation would then allow improving the current estimate of the BAO scale. Other possible applications of our semi-analytic methods are the investigation of the UV background fluctuations, which are coupled to fluctuations in the mean free path of ionizing photons on large scales (Gontcho A Gontcho et al., 2014, Pontzen, 2014), the measurement of the 3DPS of the Ly $\alpha$  forest flux, which can be exploited to improve cosmological constraints and/or constrain IGM thermal properties (McQuinn et al., 2011, McQuinn & White, 2011), and modeling the cross-correlation between Ly $\alpha$  forest and HI 21 cm signal (Guha Sarkar & Datta, 2015), as well as between CMB lensing and Ly $\alpha$  forest (Vallinotto et al., 2009, 2011).

In Chapter 3, I compared state-of-the-art hydrodynamic cosmological simulations with measurements of Ly $\alpha$  absorption both in the IGM and in the CGM, with the aim of investigating the physical properties of such media, as well as testing the predictions of simulations on a wide range of scales. More precisely, we<sup>2</sup> considered observations of the mean Ly $\alpha$  flux profile, and Ly $\alpha$  optical depth profile, around foreground QSOs, LBGs and DLAs at very different impact parameters (transverse separations) from background QSOs, ranging from 25 kpc out to 17 Mpc.

For the first time, we computed the mean Ly $\alpha$  flux profile from BOSS measurements of Ly $\alpha$ -QSO and Ly $\alpha$ -DLA cross-correlation (Font-Ribera et al. 2013 and Font-Ribera et al. 2012b, respectively), covering impact parameters in the range (1, 17) Mpc from the foreground object. We showed that these measurements extend to larger impact parameters the observations of the mean Ly $\alpha$  flux at transverse separations  $< 1$  Mpc by Prochaska et al. (2013) and Rubin et al. (2015), respectively. We compared all

---

<sup>2</sup>I am referring to the second project that I undertook during my PhD, led by myself, and done together with Dr. José Oñorbe, Prof. Dr. Joseph F. Hennawi and Dr. Zarija Lukić (Sorini et al., 2017, to be submitted).

these datasets with the predictions of Nyx (Almgren et al., 2013, Lukić et al., 2015) and Illustris (Vogelsberger et al., 2014b) hydrodynamic cosmological simulations. The former is a reference, no-feedback simulation, whereas the latter is endowed with stellar and AGN feedback prescriptions. Hereafter, I discuss the results and main conclusions distinguishing three different ranges of impact parameters.

**Large separations** ( $2 \text{ Mpc} < b < 17 \text{ Mpc}$ ) Illustris and Nyx generally reproduce well BOSS data, meaning that they both give an accurate description of the IGM (see Figures 3.4 and 3.5). Although this could be expected, it needed to be properly checked, especially considering the exquisite precision of BOSS data. For this reason, and because of the large dynamic range spanned by the observations, reproducing the mean Ly $\alpha$  flux profile at the impact parameters probed by BOSS can be used as a new fundamental test to validate cosmological simulations.

**Intermediate separations** ( $b < 2 \text{ Mpc}$ , but larger than the virial radius) Illustris and Nyx give different predictions of the mean Ly $\alpha$  flux. Such differences are mainly driven by the diverse temperature-density relationships in the CGM, stemming from the different physics implemented in the two simulations. Thus, the mean Ly $\alpha$  flux profile proves to be an excellent probe of the physics of the CGM. Thanks to their high precision, BOSS data can already constrain simulations, being able to discriminate between Nyx and Illustris at  $b \sim 1 \text{ Mpc}$  (with Illustris matching the measurement). Between the virial radius and  $\sim 400 \text{ kpc}$ , both simulations are consistent with Prochaska et al. (2013) and Rubin et al. (2015) within the error bars. We thus call for future observations focusing on increasing the statistics of background QSO - foreground QSO/DLA pairs, in order to reduce the error bars. That would enable setting tight constraints on feedback prescriptions in simulations.

**Small separations (within the virial radius)** Both Illustris and Nyx under-predict the observations, meaning that these simulations do not produce enough Ly $\alpha$  absorption in this range of impact parameters (see Figures 3.8 and 3.9). To investigate how the discrepancies could be mitigated, we developed a novel semi-analytic technique to alter the temperature-density relationship within the CGM, and applied it to the halos that we took from Nyx to reproduce the observations considered. As a result, we showed that the discrepancy between simulations and observations within the virial radius would be mitigated if simulations produced a colder CGM. However, this is not the only possibility. We argue that sub-resolution physics or, to a lesser extent, the effect of velocity fields, may partially be responsible for the discrepancy.

We showed that the different temperature and density profiles around Nyx and Illustris QSO-hosting and DLA-hosting halos are a major driver for the different predictions

of the mean Ly $\alpha$  flux at small and intermediate separations. In particular, we argue that the differences between the predictions of Nyx and Illustris out to  $\sim 3r_{\text{vir}}$  can be ascribed to the AGN feedback prescription of Illustris (see the discussion in § 3.4.1.2). Indeed, the hot component of the CGM in Illustris extends out to several virial radii from the center of the halos (Figures 3.8 and 3.9). This is the main reason why Illustris tends to systematically predict less absorption than Nyx at intermediate separations. Furthermore, injecting less heat through a weaker AGN feedback prescription should mitigate the discrepancy between Illustris and the Ly $\alpha$  absorption data within the virial radius.

Beside the temperature and density profiles, we showed that the different predictions of the mean Ly $\alpha$  flux profiles given by Nyx and Illustris can be easily interpreted in terms of the temperature-density relationship of the gas within different radial bins from the center of the halos (Figures 3.10 and 3.11). Also these plots support the claim that the effects of the Illustris feedback prescriptions extend out to  $\sim 3r_{\text{vir}}$ . We point out that such temperature-density diagrams, as well as the temperature and density profiles, are a valuable tool to get insight into the feedback prescriptions implemented in simulations, which in turn impact the Ly $\alpha$  absorption profiles. Thus, we reiterate that the Ly $\alpha$  flux profile provides a new fundamental observable to be matched when testing feedback prescriptions, complementary to other widely used quantities, such as the star formation efficiency (e.g. Behroozi et al., 2013, Guo et al., 2011, Moster et al., 2013).

As previously mentioned, the constraining power of the test presented in this thesis would benefit from the detection of a larger number of background QSO - foreground galaxy pairs. The reason is that an increase of the precision of the measurements at intermediate separations from background QSOs would allow us to put even tighter constraints on simulations. The high-precision data of Ly $\alpha$  absorption around LBGs that will be soon provided by the ongoing CLAMATO survey (Lee et al., 2014) will be useful for this purpose. From the numerical point of view, it would be interesting to repeat the comparison done in Chapter 3 with the IllustrisTNG (Pillepich et al., 2017) simulation. One of the differences with its predecessor Illustris is a weaker AGN feedback; injecting less heat, that should produce a colder CGM, thus getting closer to the observations within the virial radius.

As a perspective of this work, we will run a suite of hydrodynamic simulations, with different cosmological parameters, while keeping all other physical aspects unchanged. In fact, the incredibly high precision of BOSS data at intermediate and large scales might allow setting constraints also in this respect. We will also consider exploiting such data to discriminate among different DM models, to assess the effect of primordial magnetic fields and in general to test testing any physical effect expected to impact large scales that could be somehow traced with Ly $\alpha$  absorption. Furthermore, in this work we considered mainly the Ly $\alpha$  mean flux as a function of the impact parameter, but we plan to compute also higher-order statistics, like the PDF (see also Kollmeier

et al., 2003) and the 1DPS. These statistics may indeed provide us with further valuable information to constrain the CGM and the IGM.

In the last part of the thesis, I considered an analytic technique that will be extended to model the statistical properties of the Ly $\alpha$  forest. While numerical simulations give currently the most accurate predictions of the forest and of clustering in general, analytic methods tend to be more versatile and less computationally expensive, as they allow exploring a larger parameter space. In this respect, a recent novel approach (Bartelmann et al., 2016, 2017) based on the formalism of KFT (Das & Mazenko, 2012, 2013, Mazenko, 2010, 2011) provided the first, fully analytic prediction of the dark matter density power spectrum up to  $k \sim 10 \text{ s cMpc}^{-1}$ , matching within  $\sim 10\%$  the results of the Coyote Universe cosmological emulator (Heitmann et al. 2009, 2014, 2010, Lawrence et al. 2010; but see also the discussion in § 4.1). Given the promising results, this method is being generalized in order to predict other important quantities that can be obtained from galaxy surveys, specifically the power spectra of the kinetic energy density fluctuations of DM, as well as of the divergence and of the curl of the DM momentum density fluctuations (Littek, Sorini et al., in prep.). This study, together with the inclusion of baryons in the model (Lilow et al., in prep.), is a necessary intermediate step towards the extension of the theory to predict the statistics of the Ly $\alpha$  forest (Sorini et al., in prep.), which is the ultimate goal of my work.

In Chapter 4, we<sup>3</sup> computed the aforementioned power spectra from the Millennium Run simulation (Springel et al., 2005b, see Figure 4.3). These results will be needed to test the performance of KFT, once that is fully implemented. Meanwhile, to test the robustness of our codes, we verified that the power spectrum of density fluctuations that we computed from the Millennium Run matches the one published by Springel et al. 2005b (see Figure 4.1). We also compared it to the predictions given by KFT, showing that the agreement is generally within  $\sim 10\%$  (see also the discussion in § 4.4). The overall agreement reproduces the results found by Bartelmann et al. (2016) (see Figure 4.2).

An immediate perspective of this work is comparing the power spectra of the kinetic energy density fluctuations of DM, and of the divergence and curl of DM momentum density fluctuations, obtained from the Millennium Run, with the same quantities predicted by KFT. The size of the Millennium Run limits our comparison to  $k \gtrsim 0.0126 \text{ h cMpc}^{-1}$ . We plan to compute the aforementioned power spectra also from the Millennium XXL simulation (Angulo et al., 2012), which will allow us to probe  $k \gtrsim 0.0015 \text{ h cMpc}^{-1}$ , thus expanding the dynamic range within which we can test the predictions of KFT. Once

---

<sup>3</sup>In the last part of my PhD, I started a project, supervised by Prof. Dr. Matthias Bartelmann, aimed at predicting the statistics of the Ly $\alpha$  forest from a generalization of the KFT-based approach by Bartelmann et al. (2016, 2017). As an intermediate step towards this goal, I took part in the project led by Carsten Littek and supervised by Prof. Dr. Matthias Bartelmann, finalized at better understanding the modeling of the velocity field within KFT. The project is still ongoing, and I will be the second author on publication that will result from it (Littek, Sorini et al., in prep.). In Chapter 4 I presented my personal contribution to this particular project.

the agreement with simulations is verified, we can consider the treatment of the velocity field within the KFT approach to be accurate. Hence, at that point we will extend it to model the statistical properties of the Ly $\alpha$  forest as well.

To summarize, the general conclusion of my thesis is that the Ly $\alpha$  absorption line is destined to become a major tool to constrain both the large-scale structure of the Universe and galaxy formation. The reason is the exquisite precision of data given by recent surveys like BOSS, together with the quality of current state-of-the-art cosmological simulations, such as Illustris. Thanks to the work presented in this thesis, it is now possible to obtain accurate predictions of the statistics of the Ly $\alpha$  forest at large scales with N-body simulations, with simpler semi-analytic techniques than in the past. Furthermore, I provided a new fundamental test to validate hydrodynamic cosmological simulations, comparing them with Ly $\alpha$  absorption data in the IGM and the CGM. Such test should become standard practice for future numerical works, making it possible to provide valuable constraints for cosmology, large-scale structure and galaxy formation.

The increasingly high precision of current and upcoming observations demands a continuous improvement of theoretical modeling, in terms of accuracy, efficiency and versatility. This thesis represents an effort in this respect, and sets the stage for a more comprehensive and accurate comparison of theoretical models with observations of large-scale structure and Ly $\alpha$  absorption in the near future.



## Appendix A

### Optimal Smoothing Length for Velocity

The baryon density field can be mocked through a Gaussian smoothing of the DM density field. Also the velocity field of DM should be smoothed accordingly to reproduce the velocity field of baryons. In principle, there might be two different optimal values of  $\lambda_G$  for density and velocity, so one should vary  $\lambda_G$  for both quantities and explore all possible combinations within the dynamic range considered. However, this extensive study can be quite time consuming and is probably not the most efficient way to proceed. In all techniques tested, we shall vary  $\lambda_G$  for the density, keeping it fixed for the velocity. We determine the optimal fixed smoothing length for the velocity field as follows. We apply the FGPA (equations (2.6) and (2.1)) using the baryon density fluctuations given by our reference hydrodynamic simulation, but smoothing the DM velocity field from the same simulation at different values of  $\lambda_G$ . We then choose the smoothing length best matching the flux 1DPS, PDF and 3DPS given by the hydrodynamic simulation.

The outcome of our analysis can be seen in Figure A.1. In the left panel, we plot the accuracy in reproducing the 1DPS of the hydrodynamic simulation, versus the smoothing length. The right panel displays the analogous plot for the PDF. We notice that the smoothing of the velocities has a strong impact on the 1DPS, the optimal value being 171 ckpc, for which the mean accuracy is  $\sim 1\%$ . The PDF is less sensitive to the smoothing length for  $\lambda_G \gtrsim 285$  ckpc.

In Figure A.2 we show the results for the 3DPS. We see that the impact of the smoothing length is more important for larger  $\mu$ , whereas  $\lambda_G \lesssim 228$  ckpc yield a better accuracy with respect to  $\lambda_G \gtrsim 228$  ckpc. Given the different trends of the 1DPS, 3DPS and PDF accuracy, there is no unique optimal value of  $\lambda_G$  to maximize the accuracy in all statistics, so we have to make a compromise. We chose  $\lambda_G = 228$  ckpc/ $h$ , for which the accuracy of both PS and PDF is  $\sim 2\%$  and the accuracy of the 3DPS is between 3% and 5%, depending on the  $\mu$ -bin considered. We kept this value fixed in all our work.

Our choice of optimizing the smoothing length of the velocity allows us to focus our analysis on the impact of the smoothing length of the density field on the accuracy of our methods (see section 2.4.2 for details). As a consequence, the errors quoted for the techniques considered are minimized. Indeed, when we show the error for  $\lambda_G < 228$  ckpc

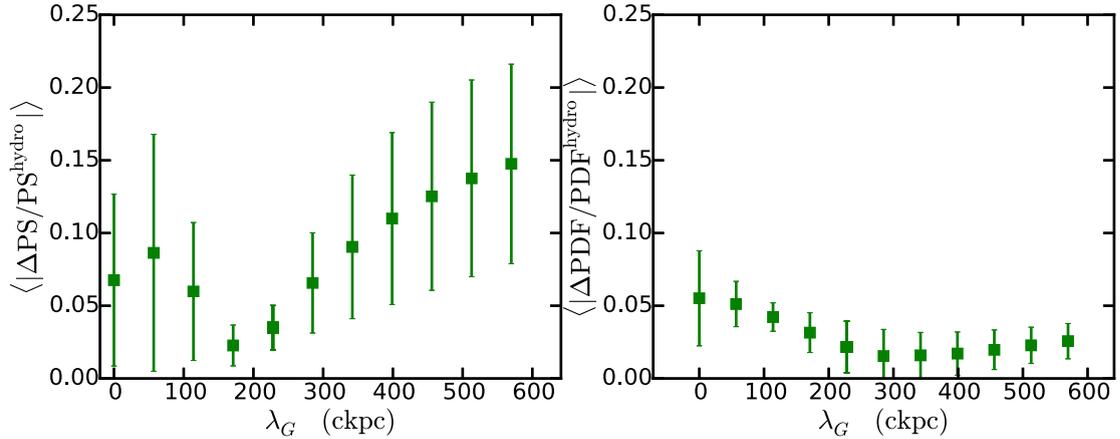


FIGURE A.1: Left panel shows the relative error between the dimensionless line-of-sight power spectrum of the flux given by the reference hydrodynamic simulation and of the flux computed by applying a 1-to-1 temperature-density relationship to the baryon density field and using the Gaussian-smoothed line-of-sight velocities of dark matter instead of baryons, as a function of different smoothing lengths of the DM velocity field. Squares mark the mean values of the accuracy, while error bars represent the root-mean-square of the accuracy in the dynamic ranges considered. The plot in the right panel is analogous, but it refers to the accuracy in reproducing the PDF of the flux in redshift space.

in the density field we are still smoothing the velocity at 228 ckpc. We recall that the smoothing length has to be at least as large as the inter-particle separation of the simulation adopted. If such separation is 228 ckpc, one can smooth the velocity field at this value and still adopted a larger  $\lambda_G$  for the DM density. Conversely, if the inter-particle separation is smaller than 228 ckpc, one can smooth the DM density choosing  $\lambda_G$  to be equal to such separation. One can still smooth the velocity field at 228 ckpc, without losing much information. Indeed, the velocity field in voids (which are the most important regions in terms of the flux statistics due to the exponentiation of equation (2.7)) is very smooth for very different cosmological models (Aragon-Calvo & Szalay, 2013, van de Weygaert & van Kampen, 1993). As such, there is no conspicuous small-scale structure in the velocity field (see also Figure 2.2), which is thus less affected by the smoothing than the DM density field. So, it is sensible to consider a fixed smoothing length for the velocity field but varying it for the DM density.

To get a sense of the accuracy obtained with a certain method adopting a different smoothing length for the velocity field, one can sum in quadrature of the errors reported in Figure 2.9 and 2.10 with the errors shown in the corresponding plots in this section, i.e. Figure A.1 and A.2. Doing so, the mean accuracy of our methods would decrease, but the trend of the accuracy versus the smoothing length would be unaffected. In particular, except for the bin along the line-of-sight of the 3DPS, GS+FGPA would still look worse than our methods.

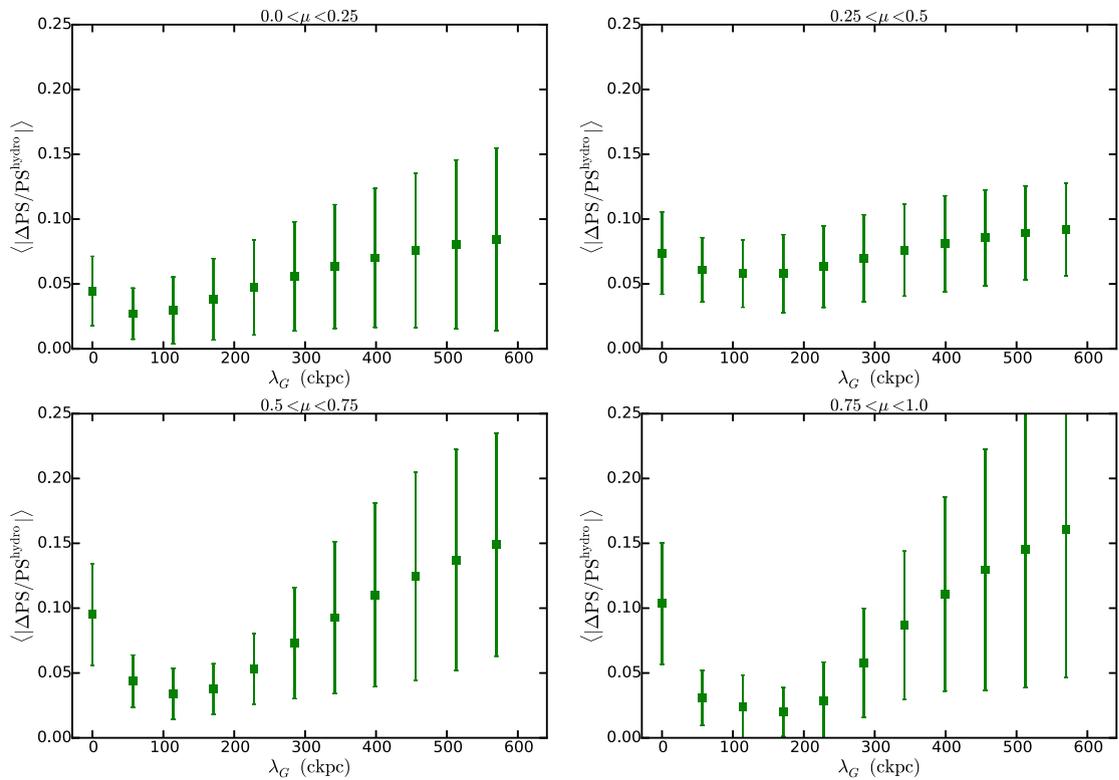


FIGURE A.2: Relative error between the dimensionless 3D power spectrum  $\Delta^2(k, \mu)$  of the flux given by the reference hydrodynamic simulation and of the flux computed by applying a 1-to-1 temperature-density relationship to the baryon density field and using the Gaussian-smoothed line-of-sight velocities of dark matter instead of baryons, as a function of different smoothing lengths of the DM velocity field. Squares mark the mean values of the accuracy, while error bars represent the root-mean-square of the accuracy in the dynamic ranges considered. Each panel refers to a different  $\mu$ -bin.

## Appendix B

### Calibration of Halo Masses

In this work we are interested in studying the physics of the CGM around QSOs, LBGs and DLAs. Such objects reside in halos with different mass. So, we can simply select halos above a certain mass threshold  $M_{\min}$ , which depends on the nature of the foreground object considered. In principle, the halo masses in Nyx and Illustris may not be calibrated in the same way, so we cannot assume the same mass threshold for both simulations.

When creating mock samples of foreground QSOs, we determine  $M_{\min}$  for Illustris and Nyx such that the resulting sample of halos reproduces the 3D two-point correlation function of quasar-hosting halos measured by White et al. (2012).

For each simulation, we select all halos such that  $M_{\text{halo}} > M_{\text{th}}$ , where  $M_{\text{halo}}$  is the halo mass as reported in the halo catalog of that simulation and  $M_{\text{th}}$  is an arbitrary threshold, which is fixed in the first place. We then compute the 3D two-point correlation function of the halos into 20 equally extended logarithmic bins of distance. We compare the correlation function with the measurements by White et al. (2012) and calculate the  $\chi^2$ . We repeat the steps just described increasing  $M_{\text{th}}$  by 0.1 dex, until there is no halo with mass larger than  $M_{\text{th}}$ . At this point, we are left with a family of correlation functions depending on  $M_{\text{th}}$ , and look for the value of the  $M_{\text{th}}$  minimizing the  $\chi^2$ . This value is the minimum mass  $M_{\min}$  that we need to set when we select halos from that simulation, such that their 3D correlation function will be as close as possible to the one observed by White et al. (2012). In other words, we assume a step function halo occupation distribution, i.e. the number of QSOs per halo is zero below  $M_{\min}$  and unity above such threshold.

The top panels of Figure B.1 illustrate the procedure for the calibration of  $M_{\min}$  for quasar hosts. In the top-left panel, the correlation function given by all Nyx halos with  $M > M_{\text{th}}$  from the snapshot at redshift  $z = 2.4$  is represented by circles, color-coded according to the corresponding value of  $M_{\text{th}}$ . The black dotted line is the analytic fit to the measurements by White et al. (2012), and the black solid line the correlation function computed from the Nyx halos, corresponding to the value of  $M_{\text{th}}$  minimizing the  $\chi^2$ . The top-right panel shows exactly the same quantities for the snapshot at

$z = 2.44$  of the Illustris simulation. We obtained  $M_{\min} = 10^{12.5} M_{\odot}$  for Nyx and  $M_{\min} = 10^{12.4} M_{\odot}$  for Illustris, in agreement with the observations by Prochaska et al. (2013). The samples of halos selected according to these thresholds have been utilized to reproduce the measurements by Font-Ribera et al. (2013) and Prochaska et al. (2013).

To simulate a sample of foreground LBGs, we apply the same procedure described for the QSOs. Instead of using observations of QSO clustering, we determine the  $M_{\min}$  that gives the best match with the observations of the 3D two-point correlation function of LBGs in the redshift range  $2 < z < 4$ , by Bielby et al. (2011).

The middle panels of Figure B.1 are analogous to the top panels, except that the black dotted lines now represent the analytic fit to the measurements by Bielby et al. (2011). We obtained  $M_{\min} = 10^{11.6} M_{\odot}$  for the  $z = 2.44$  Illustris snapshot and  $M_{\min} = 10^{11.7} M_{\odot}$  for the  $z = 2.4$  Nyx snapshot. These values of  $M_{\min}$  will be used to select the halos for the comparison of the simulations with the observations by Turner et al. (2014), who considered a sample of LBGs with  $z = 2.4$  as median redshift.

We shall also compare the simulations with the data by Adelberger et al. (2003), Adelberger et al. (2005a) and Crighton et al. (2011), who measured the radial mean flux profile around LBGs at redshifts  $z \approx 3$ ,  $z \approx 2.5$  and  $z \approx 3$ , respectively. Despite not being centered at the same redshift, the mean flux profiles measured by such observations are all normalized to  $z = 3$ . Therefore, to compare them with the simulations, we consider the Nyx snapshot at  $z = 3$  and the Illustris snapshot  $z = 3.01$ . We determined the value of  $M_{\min}$  for such snapshots, obtaining  $10^{11.5} M_{\odot}$  in both cases. The corresponding best fits to the correlation function measured by Bielby et al. (2011) are reported in the bottom panels of Figure B.1.

We notice that the values of  $M_{\min}$  inferred for our simulated LBG samples are consistent with the typical mass of LBG-hosting halos,  $\sim 10^{12} M_{\odot}$ , deduced by various authors (Adelberger et al., 2005b, Conroy et al., 2008, Rakic et al., 2013, Trainor & Steidel, 2012, Turner et al., 2014) for the KBSS survey. As discussed in § 3.2.2, we used the same halos selected for the foreground LBGs, and assumed that the DLAs lie at the center of the halos. The values of  $M_{\min}$  obtained for the LBGs are also of the same order of the characteristic mass of DLAs estimated by Font-Ribera et al. 2012b ( $10^{12} M_{\odot}$ ) from BOSS quasar spectra.

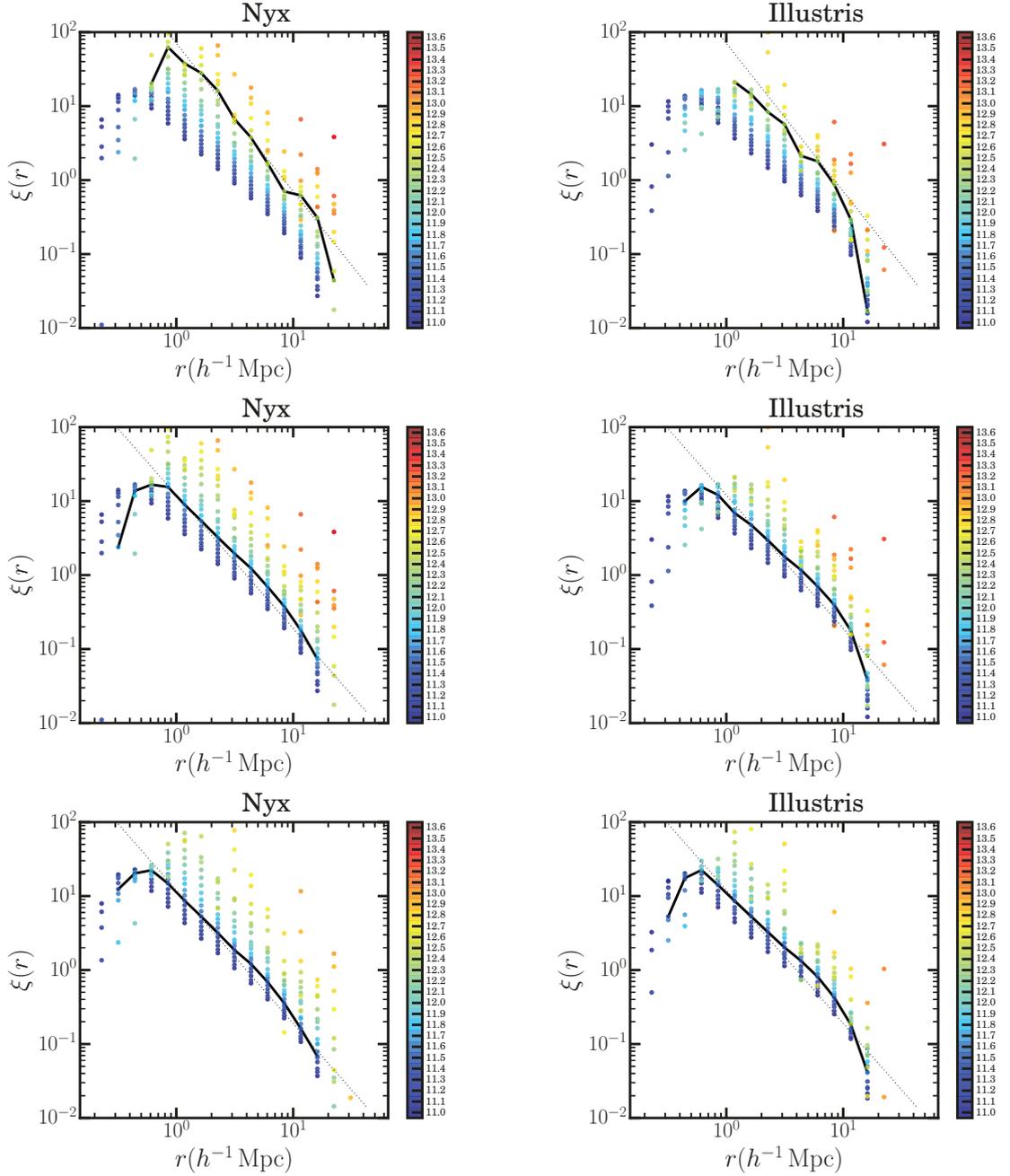


FIGURE B.1: *Top panels:* Correlation function of halos in Nyx (left panel) and Illustris (right panel), at redshift  $z = 2.4$  and  $z = 2.44$ , respectively. Dots of the same color represent the correlation function of halos with mass larger than the one corresponding to that color. The black dotted line is the analytic fit to the correlation function of quasars measured by White et al. (2012). The black solid line is the correlation function of halos above the mass threshold that best fits those observations (see text for details). *Middle panels:* As top panels, except that the black dotted line is the analytic fit to the correlation function of galaxies measured by Bielby et al. (2011), and the black solid line is the correlation function of halos above the mass threshold that best fits these observations (see text for details). *Bottom panels:* As middle panels, except that the correlation function of halos refers to the Nyx and Illustris snapshots at redshift  $z = 3$  and  $z = 3.01$ , respectively (see text for details).

## Appendix C

### Generating Mock Spectra from a Moving-Mesh Code

We simulate the absorption spectra extracting skewers on a regular grid. Since in Nyx a gas element is a cell of a Cartesian grid, the cell size of the skewers is simply given by the cell size of the simulation. Instead, Illustris treats gas on a moving mesh, constructed with a Voronoi tessellation. When we draw skewers from Illustris, we bin the gas cells into a regular grid. For this purpose, we treated each gas element as an SPH particle (following Bird et al., 2014). The smoothing length of the SPH kernel for a certain gas cell is chosen to be the maximum radius of all Delaunay tetrahedra with that cell at a vertex (see Springel, 2010a, for more details). We construct in this way the gas density, temperature and 3D velocity fields.

We need to choose the cell size of our grid so that we can have reliable predictions of the mean Ly $\alpha$  absorption around galaxies, which is the main goal of the paper. If the grid is too coarse, we may not be able to resolve the small-scale density and temperature fluctuations of the CGM. For example, if the cell size is as big as the typical virial radius, the CGM would be represented as a uniform gas cell with the average temperature and density of the CGM.

We choose a cell size equal to the mean separation of the gas cells in Illustris (58.5 ckpc), corresponding to a  $1820^3$  grid. We verified that, with a  $3550^3$  grid, corresponding to a cell size of 30 ckpc, the predicted mean Ly $\alpha$  flux fluctuations within the virial radius of QSOs and DLAs (see § 3.3) change only by 1.5%. Outside the virial radius, the difference with respect to the predictions obtained on a  $1820^3$  grid is even smaller. Since using the finest grid does not change the main results of our work, we show all predictions given by Illustris utilizing the  $1820^3$  grid. Regarding Nyx, the run that we are using has already the finest currently available grid.

## Appendix D

### From Cross-Correlation to Mean Flux Profile

The cross cross-correlation of the Ly $\alpha$  forest with DLAs (QSOs), measured by Font-Ribera et al. 2012b (Font-Ribera et al., 2013), is equivalent to the stacked mean Ly $\alpha$  flux profile around DLAs (QSOs), measured by Rubin et al. 2015 (Prochaska et al., 2013). As such, we can simply integrate the cross-correlation measurements within appropriate velocity windows along the line-of-sight, to translate such measurements into  $\delta_F$  profiles. In this section, we explain the details of the conversion done for Font-Ribera et al. (2012b) DLA-Ly $\alpha$  cross-correlation measurements. Anyway, the formalism is exactly the same also as far as the Font-Ribera et al. (2013) QSO-Ly $\alpha$  cross-correlation observations are concerned.

Font-Ribera et al. (2012b) selected a sample of 52449 QSO spectra in the redshift range  $2.1 < z < 3.5$ , from the BOSS Data Release 9. In addition, they considered a sample of 10512 foreground DLAs along the line of sights of the QSO sample, taken from the catalogue by Noterdaeme et al. (2012). Hereafter, we summarize how they estimated the cross-correlation between the DLAs and the Ly $\alpha$  absorption in the observed QSO spectra.

Font-Ribera et al. (2012b) measured the Ly $\alpha$  flux fluctuation at every pixel in each QSO spectrum. For the pixel  $i$ , it is defined as

$$\tilde{\delta}_{F_i} = \frac{f_i}{C_i \bar{F}(z_i)} - 1, \quad (\text{D.1})$$

where  $f_i$  is the measured flux,  $C_i$  is the QSO continuum, and  $\bar{F}(z_i)$  is the mean transmitted flux obtained in the redshift bin containing the pixel redshift  $z_i$ . For each DLA, the pixels of all spectra are divided into bins of transverse and line-of-sight separation from the DLA ( $b$  and  $x$ , respectively). The estimator of the cross-correlation in the bin ( $b, x$ ) is defined as

$$\hat{\xi}_{b,x} = \frac{\sum_{n=1}^N \sum_{k \in (b_n, x_n)} w_{nk} \tilde{\delta}_{F_{nk}}}{\sum_{n=1}^N \sum_{k \in (b_n, x_n)} w_{nk}}. \quad (\text{D.2})$$

The index  $n$  identifies the DLA, and  $N$  is the total number of DLAs. The index  $k$  identifies the pixels within the distance bin ( $b_n, x_n$ ) from the DLA  $n$ . The weights  $w_{nk}$

are defined as

$$w_{nk} = \left[ \sigma_F^2(z_{nk}) + \frac{\langle N_{nk}^2 \rangle}{C_{nk}^2 \bar{F}_e^2(z_{nk})} \right]^{-1}. \quad (\text{D.3})$$

In the equation above,  $\sigma_F^2(z_{nk})$  and  $\bar{F}_e^2(z_{nk})$  are the intrinsic variance of the Ly $\alpha$  forest flux fluctuations and the mean flux in the Ly $\alpha$  forest at the pixel redshift  $z_{nk}$ , respectively. For the former, Font-Ribera et al. (2012b) adopt an analytic expression based on the redshift evolution of the power spectrum measured in McDonald et al. (2006), while for the latter the observations by Faucher-Giguère et al. (2008b). Finally, the term  $\langle N_{nk}^2 \rangle$  is the noise at pixel  $nk$ , approximated as a Gaussian variance.

In Font-Ribera et al. (2012b), the  $b$ -bins are delimited by (1, 4, 7, 10, 15, 20, 30, 40, 60)  $h^{-1}$  cMpc, while the  $x$ -bins are bounded by (60, 40, 30, 20, 15, 10, 6, 3, 0)  $h^{-1}$  cMpc and the same positive values. We are interested in writing the expression of the cross-correlation  $\hat{\xi}_{b, \Delta v}$  for a transverse distance bin  $b$  and a velocity window  $\Delta v$ , centered around the DLAs. If the velocity window contains  $M$  line-of-sight distance bins as chosen by Font-Ribera et al. (2012b), equation (D.2) can be re-written as

$$\hat{\xi}_{b, \Delta v} = \frac{\sum_{n=1}^N \sum_{m=1}^M \sum_{k \in (b_n, x_{mn})} w_{nk} \tilde{\delta}_{F_{nk}}}{\sum_{n=1}^N \sum_{m=1}^M \sum_{k \in (b_n, x_{mn})} w_{nk}} = \frac{\sum_{m=1}^M \hat{\xi}_{b, x_m} \sum_{n=1}^N \sum_{k \in (b_n, x_{mn})} w_{nk}}{\sum_{n=1}^N \sum_{m=1}^M \sum_{k \in (b_n, x_{mn})} w_{nk}}. \quad (\text{D.4})$$

To compute  $\hat{\xi}_{b, \Delta v}$ , one would need to have access to all spectra, in order to properly compute the weight function at each pixel. Since we do not have access to such data, we assume that the weight function is a constant. Physically, this is equivalent to assuming that the noise term in (D.3) is the same for all pixels and that the intrinsic variance of the Ly $\alpha$  forest is approximately constant in the redshift range considered. Within such approximation, we can write

$$\hat{\xi}_{b, \Delta v} \approx \frac{1}{M} \sum_{m=1}^M K_m \hat{\xi}_{b, x_m}, \quad (\text{D.5})$$

where  $K_m$  is the number of pixels in each bin  $(b_n, x_{mn})$ . On the other hand, within the same approximation that led to (D.4), (D.2) becomes:

$$\hat{\xi}_{b, \Delta v} \approx -\delta_F(b, \Delta v), \quad (\text{D.6})$$

where  $\delta_F(b, \Delta v)$  is the mean Ly $\alpha$  flux fluctuation at impact parameter  $b$  within a velocity window  $\Delta v$  around a foreground DLA - see the definition in (3.2). Therefore, comparing (D.5) with (D.6), we obtain

$$\delta_F(b, \Delta v) \approx -\hat{\xi}_{b, \Delta v} \approx -\frac{1}{M} \sum_{m=1}^M K_m \hat{\xi}_{b, x_m}, \quad (\text{D.7})$$

within the aforementioned assumption that the weights in (D.2) are constant.

We used (D.7) to convert Font-Ribera et al. (2012b) observations into Rubin et al. (2015) measurements. The velocity window is  $\Delta v = 1000 \text{ km s}^{-1}$  (i.e.  $\pm 500 \text{ km s}^{-1}$  around the foreground object), which corresponds to  $\sim 10 h^{-1} \text{ cMpc}$  assuming the same cosmology as Font-Ribera et al. (2012b). The results of our analysis are listed in Table D.1, where we determined the errors propagating the errors in Font-Ribera et al. (2012b) measurements. As a caveat, we point out that our estimate (D.7) would be exact if Font-Ribera et al. (2012b) data were re-analyzed computing the weights defined in (D.3) pixel by pixel.

The comparison between Prochaska et al. (2013) and Font-Ribera et al. (2013) can be done following the same argument explained in this appendix, using a velocity window of  $2000 \text{ km s}^{-1}$ . In this case, the line-of-sight bins chosen by Font-Ribera et al. (2013) do not cover exactly the desired velocity window, so we linearly interpolate between their data points in order to integrate them within Prochaska et al. (2013) velocity window. The results are tabulated in table D.2. It is the first time that large-scale measurements of the Ly $\alpha$  cross-correlation function (Font-Ribera et al., 2013, 2012b) are used together with observations of Ly $\alpha$  absorption in the CGM (Prochaska et al., 2013, Rubin et al., 2015) to jointly constrain the physics of the IGM and the CGM.

$b_{\min}^a$ ( $h^{-1}$ cMpc)	$b_{\max}^b$ ( $h^{-1}$ cMpc)	$\delta_F^c$
1	4	$0.082 \pm 0.012$
4	7	$0.0509 \pm 0.0080$
7	10	$0.0521 \pm 0.0070$
10	15	$0.0271 \pm 0.0045$
15	20	$0.0183 \pm 0.0038$
20	30	$0.0104 \pm 0.0023$
30	40	$0.0056 \pm 0.0019$
40	60	$0.0027 \pm 0.0012$

<sup>a</sup> Inner edge of the impact parameter bin.

<sup>b</sup> Outer edge of the impact parameter bin.

<sup>c</sup> Mean Ly $\alpha$  flux fluctuation.

TABLE D.1: Ly $\alpha$  absorption at large impact parameter from DLAs, inferred from Font-Ribera et al. (2012b)

$b_{\min}^a$ ( $h^{-1}$ cMpc)	$b_{\max}^b$ ( $h^{-1}$ cMpc)	$\delta_F^c$
1	4	$0.0669 \pm 0.0030$
4	7	$0.0509 \pm 0.0020$
7	10	$0.0378 \pm 0.0017$
10	15	$0.0259 \pm 0.0011$
15	10	$0.01945 \pm 0.00091$
20	30	$0.01051 \pm 0.00055$
30	40	$0.00406 \pm 0.00047$
40	60	$0.00269 \pm 0.00030$

<sup>a</sup> Inner edge of the impact parameter bin.

<sup>b</sup> Outer edge of the impact parameter bin.

<sup>c</sup> Mean Ly $\alpha$  flux fluctuation.

TABLE D.2: Ly $\alpha$  absorption at large impact parameter from QSOs, inferred from Font-Ribera et al. (2013)

## Appendix E

### Mean Flux Fluctuations at Small Separations from DLAs

Rubin et al. (2015) stacked the absorption spectra of background QSOs passing at different transverse separation from foreground DLAs in four bins of impact parameter. After re-normalizing the spectra to the pseudo-continuum measured in the velocity intervals  $(-4000, -3500) \text{ km s}^{-1}$  and  $(3500, 4000) \text{ km s}^{-1}$ , they determined the equivalent width of the Ly $\alpha$  absorption  $W_{\text{Ly}\alpha}$  in the velocity window  $\Delta v = 1000 \text{ km s}^{-1}$  around the DLAs. From the definition of equivalent width (Draine, 2011), the mean Ly $\alpha$  flux  $\langle F \rangle$  centered on the DLA can be inferred as

$$\langle F \rangle = F_0 \left( 1 - \frac{c W_{\text{Ly}\alpha}}{\Delta v \lambda_{\text{Ly}\alpha}} \right), \quad (\text{E.1})$$

where  $c$  is the speed of light in vacuum,  $\lambda_{\text{Ly}\alpha}$  the rest-frame wavelength of the Ly $\alpha$  transition, and  $F_0$  the pseudo-continuum. Since the pseudo-continuum measured by Rubin et al. (2015) is meant to represent the mean flux of the IGM at the redshift of their observations, we can infer the mean Ly $\alpha$  flux fluctuations simply as

$$\delta_F = \frac{c W_{\text{Ly}\alpha}}{\Delta v \lambda_{\text{Ly}\alpha}}. \quad (\text{E.2})$$

The results are listed in Table E.1.

$b_{\min}$ <sup>a</sup> (kpc)	$b_{\max}$ <sup>b</sup> (kpc)	$\delta_F$ <sup>c</sup>
0	50	$0.436 \pm 0.092$
50	100	$0.345 \pm 0.082$
100	200	$0.269 \pm 0.010$
200	300	$0.037 \pm 0.057$

<sup>a</sup> Inner edge of the impact parameter bin.

<sup>b</sup> Outer edge of the impact parameter bin.

<sup>c</sup> Mean Ly $\alpha$  flux fluctuation.

TABLE E.1: Ly $\alpha$  absorption around DLAs obtained from Rubin et al. (2015)

## Appendix F

### Outflows in the CGM

In § 3.4.1 we focused on the impact of density and temperature of the gas on Ly $\alpha$  absorption. Nonetheless, the Ly $\alpha$  optical depth depends also on the peculiar velocity of the gas. In fact, the smoothness of the velocity field can have a significant impact on the statistics of the Ly $\alpha$  absorption lines (Sorini et al., 2016). Moreover, various hydrodynamic simulations (including Illustris) underpredict the line width distribution of the Ly $\alpha$  forest (Viel et al., 2017) or the line width - HI column density relationship (Gaikwad et al., 2017) obtained from HST-COS QSO absorption spectra. The agreement with data of the latter statistics can be improved adding a turbulent broadening contribution to the line width in the simulations (Gaikwad et al. 2017; see also Oppenheimer & Davé 2009 and Viel et al. 2017). This term is not a thermal broadening, but velocity broadening coming from motions not captured by the simulations.

In this section, we focus on the connection between the radial component of the velocity of the gas around galaxies and Ly $\alpha$  absorption in the CGM. Following what we did in § 3.4.1, we investigate the radial velocity - hydrogen density relationship of the gas in the CGM in Nyx and Illustris. In Figure F.1 we plot this relationship around  $M \gtrsim 10^{12.5} M_{\odot}$  ( $M \gtrsim 10^{12.4} M_{\odot}$ ) halos from Nyx (Illustris), within the same radial bins as in § 3.4.1. In Figure F.2 we show an analogous plot for the  $M \gtrsim 10^{11.7} M_{\odot}$  ( $M \gtrsim 10^{11.6} M_{\odot}$ ) halos from Nyx (Illustris).

Figure F.1 shows that, at any given radial bin, the shape of the radial velocity - hydrogen density relationship is qualitatively similar in Nyx and Illustris. However, the gas in Illustris presents an overall offset of  $\sim +50 \text{ km s}^{-1}$  in the radial velocity with respect to Nyx. Furthermore, in the innermost bin, Nyx presents a larger spread in radial velocity: the radial velocity of the most rapidly inflowing gas in Nyx is  $\sim -200 \text{ km s}^{-1}$ , while in Illustris we do not find gas inflowing faster than  $150 \text{ km s}^{-1}$ . In the interval  $(r_{\text{vir}}, 2r_{\text{vir}})$ , the majority of the gas in Nyx lies in the radial velocity range  $(-200, -100)$ . On the contrary, there is a larger amount of gas with positive radial velocity (i.e. outflowing) in Illustris. Finally, in the bin  $(2r_{\text{vir}}, 3r_{\text{vir}})$ , Illustris presents a plume toward more positive velocities, corresponding to gas with radial velocity  $\gtrsim -50 \text{ km s}^{-1}$  and density in the

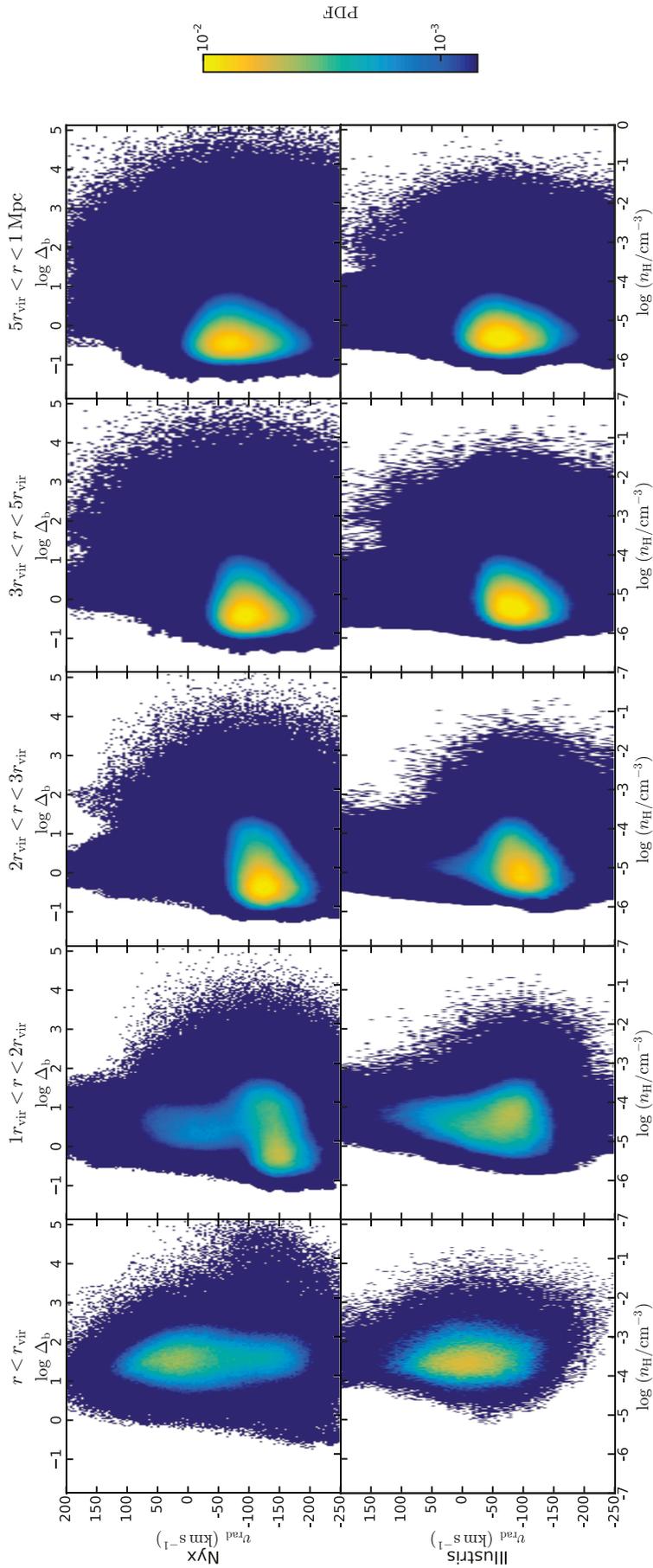


FIGURE F.1: Radial velocity - density relationship of hydrogen at different radial bins from a foreground QSO. The volume-weighted 2D histograms are plotted upon stacking 100 randomly drawn QSO-hosting halos in Nyx ( $> 10^{12.5} M_{\odot}$ , top panels) and Illustris ( $> 10^{12.4} M_{\odot}$ , bottom panels), in different bins of distance from the center of the halos. From left to right, the histograms refer to the radial bins ( $0, r_{\text{vir}}, 2r_{\text{vir}}, 3r_{\text{vir}}, 5r_{\text{vir}}$ ) and ( $5r_{\text{vir}}, 1 \text{ Mpc}$ ). Negative velocities denote inflowing gas. Overall, the two simulations yield qualitatively similar diagrams at all virial radii. Illustris presents an overall offset of  $\sim +50 \text{ km s}^{-1}$  with respect to Nyx in the three innermost bins, and generally more outflowing gas than Nyx (see text for details).

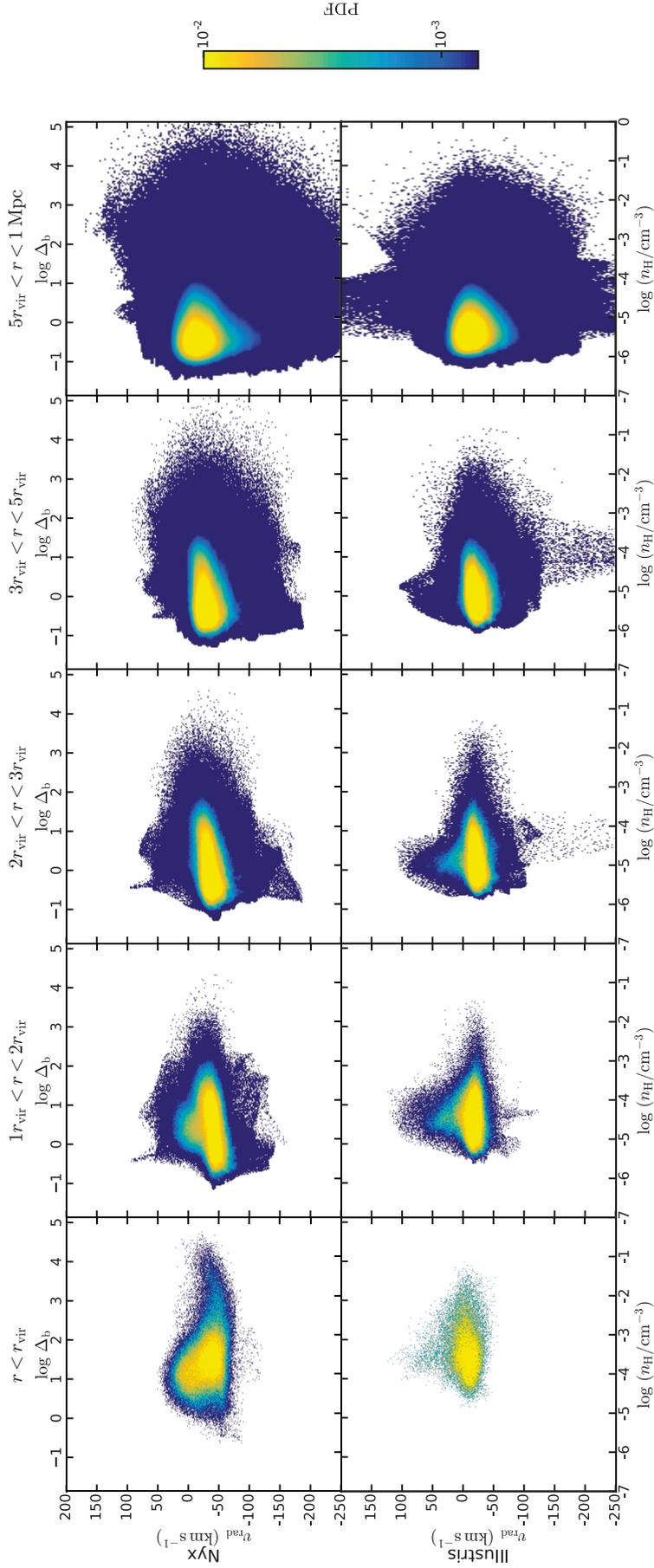


FIGURE F.2: Same as in Figure F.1, but for DLA/LBG-hosting halos in Nyx ( $> 10^{11.7} M_{\odot}$ , top panels) and Illustris ( $> 10^{11.6} M_{\odot}$ ). Also in this case, the two simulations yield overall qualitatively similar diagrams at all virial radii. Illustris presents an overall offset of  $\sim +20 \text{ km s}^{-1}$  with respect to Nyx in the three innermost bins, and generally more outflowing gas than Nyx (see text for details).

range  $(10^{-5.5}, 10^{-4.5}) \text{ cm}^{-3}$ . Such feature is absent in Nyx, instead. For  $r > 3r_{\text{vir}}$ , Nyx and Illustris present very similar radial velocity - density diagrams.

In Figure F.2, the radial velocity - hydrogen density relationships in Nyx and Illustris look qualitatively even more similar than in Figure F.1. The most different bin is the innermost one, where Nyx presents a larger spread in the diagram. In all radial bins, the gas in Illustris appears to have an overall offset of  $\sim +20 \text{ km s}^{-1}$  with respect to the gas in Nyx. In the bin  $(2r_{\text{vir}}, 3r_{\text{vir}})$ , Illustris presents an excess of outflowing gas in the density range  $(10^{-5.5}, 10^{-4.5}) \text{ cm}^{-3}$  with respect to Nyx.

Analyzing the radial velocity - density relationships of hydrogen in the CGM in the two simulations, we can conclude that, in general, Illustris presents more outflowing gas than Nyx. However, the velocity offsets are small compared to the velocity window within which  $\delta_F$  is computed ( $1000 \text{ km s}^{-1}$  and  $2000 \text{ km s}^{-1}$  for DLAs and LBGs, respectively;  $340 \text{ km s}^{-1}$  for Turner et al. 2014 measurements). For this reason, and considering that the radial velocity - density relationships in the two simulations look much more alike than the respective temperature - density relationships, we think that the higher CGM temperature in Illustris has a greater impact on the Ly $\alpha$  absorption profiles than the larger amount of outflowing gas.

## Appendix G

### Convergence of Power Spectra

As explained in § 4.4, we computed all power spectra after binning the density and momenta of the particles in the Millennium Run onto a regular Cartesian grid. In this section, we evaluate the impact of the cell size of such grid on the resulting power spectrum.

In the left panel of Figure G.1 we show the power spectrum of density fluctuations computed from a regular Cartesian grid with  $3000^3$ ,  $2160^3$ ,  $1080^3$ ,  $540^3$  and  $270^3$  cells (solid black, red, green, blue and purple lines, respectively). The shaded areas represent the scatter due to cosmic variance. The power spectra agree at small  $k$ , but present different cutoff scales at large  $k$ . This is because finer grids present a larger Nyquist mode. In the right panel of Figure G.1, with the same color coding as the left panel, we show the relative difference of each power spectrum with respect to the one computed from the finest grid. The shaded areas represent the uncertainty on the relative difference obtained propagating the errors due to cosmic variance. We notice that, for the grid with  $2160^3$  cells, the results are converged up to  $k \sim 3 h \text{cMpc}^{-1}$  within  $\sim 5\%$ . Such

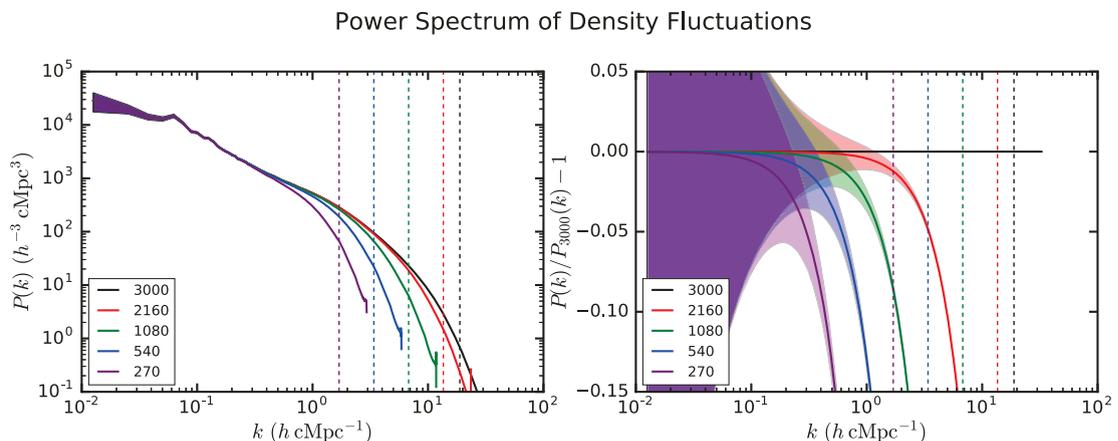


FIGURE G.1: Power spectrum of the curl of momentum overdensity at redshift  $z = 0$ , predicted by KFT (solid black line) and by the Millennium Run (solid orange line). The orange shaded area represents the error due to cosmic variance in the Millennium Run. The vertical dotted orange line marks the Nyquist mode in the grid used to compute the power spectra in the Millennium Run (see text for details).

agreement is consistent with the one obtained by Schneider et al. (2016) in an analogous test with the `Pkdgrav3` N-body code (Stadel, 2001).

In Figure G.2, from top to bottom, we show analogous plots for the power spectra of the kinetic energy density fluctuations, and of the divergence and curl of the momentum density fluctuations, respectively. We use the same color coding as in Figure G.1. We find out that the results from the grid with  $2160^3$  converge to the predictions of the grid with  $3000^3$  cells within 5% up to  $3 h cMpc^{-1}$  for all power spectra. Thus, the tests done in this section ensure that our preliminary results, obtained from the grid with  $2160^3$  cells and presented in chapter 4, are robust.

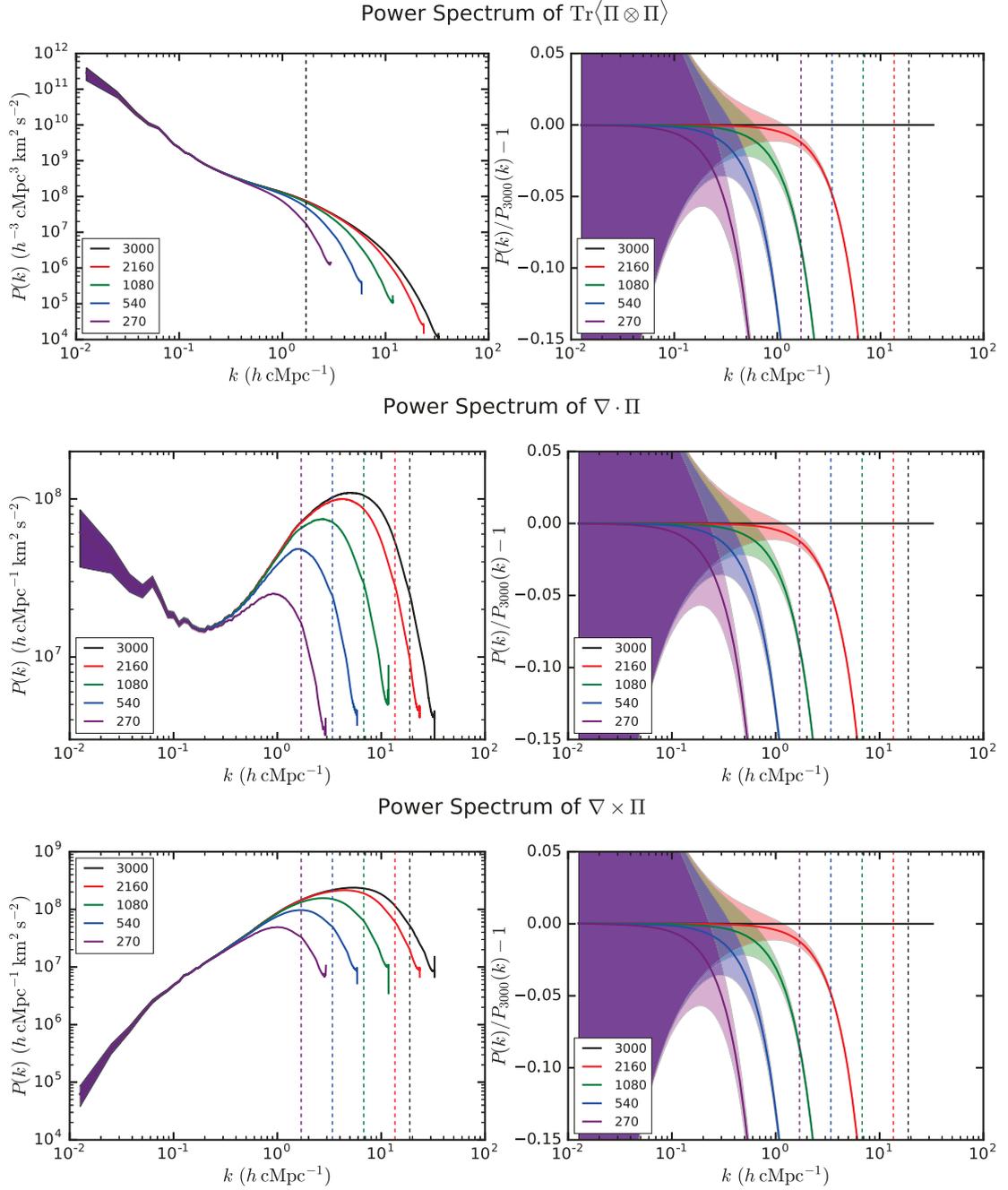


FIGURE G.2: From top to bottom, power spectra of the kinetic energy density fluctuations, of the divergence and of the curl of momentum density fluctuations, respectively, obtained from the snapshot at redshift  $z = 0$  of the Millennium Run, after CIC-binning the dark matter particles onto a regular Cartesian grid with  $3000^3$ ,  $2160^3$ ,  $1080^3$ ,  $540^3$  and  $270^3$  cells (solid black, red, green, blue and purple lines, respectively). The power spectra computed from  $2160^3$ , used to obtain the results presented in Chapter 4, are converged within 5% up to  $k \sim 3 h \text{cMpc}^{-1}$ .

## List of Author's Publications

### Published Works

1. **D. Sorini**, *An Optimally Weighted Estimator of the Linear Power Spectrum Distinguishing the Growth of Density Perturbations Across Galaxy Surveys*, JCAP, Issue 04, article id. 029 (2017)
2. **D. Sorini**, J. Oñorbe, Z. Lukić, J. F. Hennawi, *Modeling the Ly $\alpha$  Forest in Collisionless Simulations*, ApJ, Vol. 827, Nr. 2, p. 97 (2016)\*

### Unpublished Works

1. **D. Sorini**, J. Oñorbe, J. F. Hennawi, Z. Lukić, *A New Fundamental Test For Feedback Prescriptions in Cosmological Simulations: Matching Ly $\alpha$  Absorption in the Circumgalactic and Intergalactic Media*, to be submitted (2017)\*
2. **D. Sorini**, M. Bartelmann *et al.*, in prep.
3. C. Littek, **D. Sorini**, M. Bartelmann *et al.*, in prep.\*

---

\* Used in this thesis.

## Bibliography

- Abel, T., Bryan, G. L., & Norman, M. L. 2002, *Science*, 295, 93
- Adelberger, K. L., Shapley, A. E., Steidel, C. C., Pettini, M., Erb, D. K., & Reddy, N. A. 2005a, *ApJ*, 629, 636
- Adelberger, K. L., Steidel, C. C., Pettini, M., Shapley, A. E., Reddy, N. A., & Erb, D. K. 2005b, *ApJ*, 619, 697
- Adelberger, K. L., Steidel, C. C., Shapley, A. E., & Pettini, M. 2003, *ApJ*, 584, 45
- Agertz, O., Moore, B., Stadel, J., Potter, D., Miniati, F., Read, J., Mayer, L., Gawryszczak, A., Kravtsov, A., Nordlund, Å., Pearce, F., Quilis, V., Rudd, D., Springel, V., Stone, J., Tasker, E., Teyssier, R., Wadsley, J., & Walder, R. 2007, *MNRAS*, 380, 963
- Ahn, C. P., Alexandroff, R., Allende Prieto, C., Anderson, S. F., Anderton, T., Andrews, B. H., Aubourg, É., Bailey, S., Balbinot, E., Barnes, R., & et al. 2012, *ApJS*, 203, 21
- Almgren, A. S., Bell, J. B., Lijewski, M. J., Lukić, Z., & Van Andel, E. 2013, *ApJ*, 765, 39
- Amendola, L., Appleby, S., Bacon, D., Baker, T., Baldi, M., Bartolo, N., Blanchard, A., Bonvin, C., Borgani, S., Branchini, E., Burrage, C., Camera, S., Carbone, C., Casarini, L., Cropper, M., de Rham, C., Di Porto, C., Ealet, A., Ferreira, P. G., Finelli, F., García-Bellido, J., Giannantonio, T., Guzzo, L., Heavens, A., Heisenberg, L., Heymans, C., Hoekstra, H., Hollenstein, L., Holmes, R., Horst, O., Jahnke, K., Kitching, T. D., Koivisto, T., Kunz, M., La Vacca, G., March, M., Majerotto, E., Markovic, K., Marsh, D., Marulli, F., Massey, R., Mellier, Y., Mota, D. F., Nunes, N. J., Percival, W., Pettorino, V., Porciani, C., Quercellini, C., Read, J., Rinaldi, M., Sapone, D., Scaramella, R., Skordis, C., Simpson, F., Taylor, A., Thomas, S., Trotta, R., Verde, L., Vernizzi, F., Vollmer, A., Wang, Y., Weller, J., & Zlosnik, T. 2013, *Living Reviews in Relativity*, 16, 6
- Amendola, L., & Tsujikawa, S. 2010, *Dark Energy: Theory and Observations*
- Angulo, R. E., Springel, V., White, S. D. M., Jenkins, A., Baugh, C. M., & Frenk, C. S. 2012, *MNRAS*, 426, 2046
- Anselmi, S., Matarrese, S., & Pietroni, M. 2011, *JCAP*, 6, 015
- Anselmi, S., & Pietroni, M. 2012, *JCAP*, 12, 013
- Aragon-Calvo, M. A., & Szalay, A. S. 2013, *MNRAS*, 428, 3409
- Arun, K., Gudennavar, S. B., & Sivaram, C. 2017, *Advances in Space Research*, 60, 166

- Bagla, J. S. 1998, MNRAS, 299, 417
- . 2002, *Journal of Astrophysics and Astronomy*, 23, 185
- Bahcall, J. N., & Salpeter, E. E. 1965, ApJ, 142, 1677
- Baldry, I. K., Driver, S. P., Loveday, J., Taylor, E. N., Kelvin, L. S., Liske, J., Norberg, P., Robotham, A. S. G., Brough, S., Hopkins, A. M., Bamford, S. P., Peacock, J. A., Bland-Hawthorn, J., Conselice, C. J., Croom, S. M., Jones, D. H., Parkinson, H. R., Popescu, C. C., Prescott, M., Sharp, R. G., & Tuffs, R. J. 2012, MNRAS, 421, 621
- Baldry, I. K., Glazebrook, K., & Driver, S. P. 2008, MNRAS, 388, 945
- Barnes, J., & Hut, P. 1986, *Nature*, 324, 446
- Barnes, L. A., Garel, T., & Kacprzak, G. G. 2014, *Publications of the Astronomical Society of the Pacific*, 126, 969
- Barrow, J. D., & Turner, M. S. 1981, *Nature*, 292, 35
- Bartelmann, M. 2015, *Phys. Rev. D*, 91, 083524
- Bartelmann, M., Fabis, F., Berg, D., Kozlikin, E., Lilow, R., & Viermann, C. 2016, *New Journal of Physics*, 18, 043020
- Bartelmann, M., Felix Fabis, F., Kozlikin, E., Lilow, R., Dombrowski, J., & Mildemberger, J. 2017, *New Journal of Physics*, 19, 083001
- Bartelmann et al. in prep.
- Baumgart, D. J., & Fry, J. N. 1991, ApJ, 375, 25
- Bautista, J. E., Bailey, S., Font-Ribera, A., Pieri, M. M., Busca, N. G., Miralda-Escudé, J., Palanque-Delabrouille, N., Rich, J., Dawson, K., Feng, Y., Ge, J., Gontcho, S. G. A., Ho, S., Le Goff, J. M., Noterdaeme, P., Pâris, I., Rossi, G., & Schlegel, D. 2015, JCAP, 5, 060
- Becker, G. D., & Bolton, J. S. 2013, MNRAS, 436, 1023
- Becker, G. D., Hewett, P. C., Worseck, G., & Prochaska, J. X. 2013, MNRAS, 430, 2067
- Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, ApJ, 762, L31
- Belikov, A. V., & Hooper, D. 2009, *Phys. Rev. D*, 80, 035007
- Benson, A. J. 2010, *Phys. Rep.*, 495, 33
- Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. 2002, *Phys. Rep.*, 367, 1
- Bernardi, M., Meert, A., Sheth, R. K., Vikram, V., Huertas-Company, M., Mei, S., & Shankar, F. 2013, MNRAS, 436, 697
- Bielby, R. M., Shanks, T., Weilbacher, P. M., Infante, L., Crighton, N. H. M., Bornancini, C., Bouché, N., Héraudeau, P., Lambas, D. G., Lowenthal, J., Minniti, D., Padilla, N., Petitjean, P., & Theuns, T. 2011, MNRAS, 414, 2

- Binney, J., & Tremaine, S. 2008, *Galactic Dynamics: Second Edition* (Princeton University Press)
- Bird, S., Haehnelt, M., Neeleman, M., Genel, S., Vogelsberger, M., & Hernquist, L. 2015, *MNRAS*, 447, 1834
- Bird, S., Vogelsberger, M., Haehnelt, M., Sijacki, D., Genel, S., Torrey, P., Springel, V., & Hernquist, L. 2014, *MNRAS*, 445, 2313
- Bird, S., Vogelsberger, M., Sijacki, D., Zaldarriaga, M., Springel, V., & Hernquist, L. 2013, *MNRAS*, 429, 3341
- Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, *Nature*, 311, 517
- Bode, P., Ostriker, J. P., & Xu, G. 2000, *ApJS*, 128, 561
- Bolton, J. S., Puchwein, E., Sijacki, D., Haehnelt, M. G., Kim, T.-S., Meiksin, A., Regan, J. A., & Viel, M. 2017, *MNRAS*, 464, 897
- Borde, A., Palanque-Delabrouille, N., Rossi, G., Viel, M., Bolton, J. S., Yèche, C., LeGoff, J.-M., & Rich, J. 2014, *JCAP*, 7, 005
- Bouchet, F. R., Colombi, S., Hivon, E., & Juszkiewicz, R. 1995, *A&A*, 296, 575
- Bromm, V., & Yoshida, N. 2011, *ARA&A*, 49, 373
- Brown, P. N., Byrne, G. D., & Hindmarsh, A. C. 1989, *SIAM J. Sci. Stat. Comput.*, 10, 1038
- Caldwell, R. R., Dave, R., & Steinhardt, P. J. 1998, *Ap&SS*, 261, 303
- Campbell, L. A., Lucey, J. R., Colless, M., Jones, D. H., Springob, C. M., Magoulas, C., Proctor, R. N., Mould, J. R., Read, M. A., Brough, S., Jarrett, T., Merson, A. I., Lah, P., Beutler, F., Cluver, M. E., & Parker, Q. A. 2014, *MNRAS*, 443, 1231
- Cantalupo, S. 2010, *MNRAS*, 403, L16
- Cen, R., Miralda-Escudé, J., Ostriker, J. P., & Rauch, M. 1994, *ApJ*, 437, L9
- Ceverino, D., Dekel, A., & Bournaud, F. 2010, *MNRAS*, 404, 2151
- Ceverino, D., Dekel, A., Mandelker, N., Bournaud, F., Burkert, A., Genzel, R., & Primack, J. 2012, *MNRAS*, 420, 3490
- Chardin, J., Haehnelt, M. G., Aubert, D., & Puchwein, E. 2015, *MNRAS*, 453, 2943
- Clerkin, L., Kirk, D., Lahav, O., Abdalla, F. B., & Gaztañaga, E. 2015, *MNRAS*, 448, 1389
- Cole, S., Ellis, R. S., Broadhurst, T. J., & Colless, M. M. 1994, *MNRAS*, 267, 541
- Colella, P. 1990, *Journal of Computational Physics*, 87, 171
- Colombi, S., Bouchet, F. R., & Hernquist, L. 1996, *ApJ*, 465, 14
- Conroy, C., Shapley, A. E., Tinker, J. L., Santos, M. R., & Lemson, G. 2008, *ApJ*, 679, 1192

- Cooray, A., & Sheth, R. 2002, *Phys. Rep.*, 372, 1
- Couch, W. J., Jurcevic, J. S., & Boyle, B. J. 1993, *MNRAS*, 260, 241
- Crain, R. A., Bahé, Y. M., Lagos, C. d. P., Rahmati, A., Schaye, J., McCarthy, I. G., Marasco, A., Bower, R. G., Schaller, M., Theuns, T., & van der Hulst, T. 2017, *MNRAS*, 464, 4204
- Crain, R. A., Schaye, J., Bower, R. G., Furlong, M., Schaller, M., Theuns, T., Dalla Vecchia, C., Frenk, C. S., McCarthy, I. G., Helly, J. C., Jenkins, A., Rosas-Guevara, Y. M., White, S. D. M., & Trayford, J. W. 2015, *MNRAS*, 450, 1937
- Crichton, N. H. M., Bielby, R., Shanks, T., Infante, L., Bornancini, C. G., Bouché, N., Lambas, D. G., Lowenthal, J. D., Minniti, D., Morris, S. L., Padilla, N., Péroux, C., Petitjean, P., Theuns, T., Tummuangpak, P., Weibacher, P. M., Wisotzki, L., & Worseck, G. 2011, *MNRAS*, 414, 28
- Crichton, N. H. M., Hennawi, J. F., & Prochaska, J. X. 2013, *ApJ*, 776, L18
- Crichton, N. H. M., Hennawi, J. F., Simcoe, R. A., Cooksey, K. L., Murphy, M. T., Fumagalli, M., Prochaska, J. X., & Shanks, T. 2015, *MNRAS*, 446, 18
- Croft, R. A. C., Hernquist, L., Springel, V., Westover, M., & White, M. 2002, *ApJ*, 580, 634
- Croft, R. A. C., Weinberg, D. H., Katz, N., & Hernquist, L. 1998, in *Large Scale Structure: Tracks and Traces*, ed. V. Mueller, S. Gottloeber, J. P. Muecket, & J. Wambsganss, 69–75
- Croft, R. A. C., Weinberg, D. H., Pettini, M., Hernquist, L., & Katz, N. 1999, *ApJ*, 520, 1
- da Costa, L. N., Vogeley, M. S., Geller, M. J., Huchra, J. P., & Park, C. 1994, *ApJ*, 437, L1
- Das, S. P., & Mazenko, G. F. 2012, *Journal of Statistical Physics*, 149, 643
- . 2013, *Journal of Statistical Physics*, 152, 159
- Davies, F. B., & Furlanetto, S. R. 2015, *ArXiv e-prints*
- Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, *ApJ*, 292, 371
- Davis, M., & Peebles, P. J. E. 1977, *ApJS*, 34, 425
- Dawson, K. S., Schlegel, D. J., Ahn, C. P., Anderson, S. F., Aubourg, É., Bailey, S., Barkhouser, R. H., Bautista, J. E., Beifiori, A., Berlind, A. A., Bhardwaj, V., Bizyaev, D., Blake, C. H., Blanton, M. R., Blomqvist, M., Bolton, A. S., Borde, A., Bovy, J., Brandt, W. N., Brewington, H., Brinkmann, J., Brown, P. J., Brownstein, J. R., Bundy, K., Busca, N. G., Carithers, W., Carnero, A. R., Carr, M. A., Chen, Y., Comparat, J., Connolly, N., Cope, F., Croft, R. A. C., Cuesta, A. J., da Costa, L. N., Davenport, J. R. A., Delubac, T., de Putter, R., Dhital, S., Ealet, A., Ebelke, G. L., Eisenstein, D. J., Escoffier, S., Fan, X., Filiz Ak, N., Finley, H., Font-Ribera, A., Génova-Santos, R., Gunn, J. E., Guo, H., Haggard, D., Hall, P. B., Hamilton, J.-C., Harris, B., Harris, D. W., Ho, S., Hogg, D. W., Holder, D., Honscheid, K., Huehnerhoff, J., Jordan, B., Jordan, W. P., Kauffmann, G., Kazin, E. A., Kirkby, D., Klaene,

- M. A., Kneib, J.-P., Le Goff, J.-M., Lee, K.-G., Long, D. C., Loomis, C. P., Lundgren, B., Lupton, R. H., Maia, M. A. G., Makler, M., Malanushenko, E., Malanushenko, V., Mandelbaum, R., Manera, M., Maraston, C., Margala, D., Masters, K. L., McBride, C. K., McDonald, P., McGreer, I. D., McMahon, R. G., Mena, O., Miralda-Escudé, J., Montero-Dorta, A. D., Montesano, F., Muna, D., Myers, A. D., Naugle, T., Nichol, R. C., Noterdaeme, P., Nuza, S. E., Olmstead, M. D., Oravetz, A., Oravetz, D. J., Owen, R., Padmanabhan, N., Palanque-Delabrouille, N., Pan, K., Parejko, J. K., Pâris, I., Percival, W. J., Pérez-Fournon, I., Pérez-Ràfols, I., Petitjean, P., Pfaffenberger, R., Pforr, J., Pieri, M. M., Prada, F., Price-Whelan, A. M., Raddick, M. J., Rebolo, R., Rich, J., Richards, G. T., Rockosi, C. M., Roe, N. A., Ross, A. J., Ross, N. P., Rossi, G., Rubiño-Martin, J. A., Samushia, L., Sánchez, A. G., Sayres, C., Schmidt, S. J., Schneider, D. P., Scóccola, C. G., Seo, H.-J., Shelden, A., Sheldon, E., Shen, Y., Shu, Y., Slosar, A., Smee, S. A., Snedden, S. A., Stauffer, F., Steele, O., Strauss, M. A., Streblyanska, A., Suzuki, N., Swanson, M. E. C., Tal, T., Tanaka, M., Thomas, D., Tinker, J. L., Tojeiro, R., Tremonti, C. A., Vargas Magaña, M., Verde, L., Viel, M., Wake, D. A., Watson, M., Weaver, B. A., Weinberg, D. H., Weiner, B. J., West, A. A., White, M., Wood-Vasey, W. M., Yeche, C., Zehavi, I., Zhao, G.-B., & Zheng, Z. 2013, *AJ*, 145, 10
- de Jong, R. S., Barden, S. C., Bellido-Tirado, O., Brynnel, J. G., Frey, S., Giannone, D., Haynes, R., Johl, D., Phillips, D., Schnurr, O., Walcher, J. C., Winkler, R., Ansorge, W. R., Feltzing, S., McMahon, R. G., Baker, G., Caillier, P., Dwelly, T., Gaessler, W., Iwert, O., Mandel, H. G., Piskunov, N. A., Pragt, J. H., Walton, N. A., Bensby, T., Bergemann, M., Chiappini, C., Christlieb, N., Cioni, M.-R. L., Driver, S., Finoguenov, A., Helmi, A., Irwin, M. J., Kitaura, F.-S., Kneib, J.-P., Liske, J., Merloni, A., Minchev, I., Richard, J., & Starkenburg, E. 2016, in *Proc. SPIE*, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, 99081O
- Dekel, A., Zolotov, A., Tweed, D., Cacciato, M., Ceverino, D., & Primack, J. R. 2013, *MNRAS*, 435, 999
- Delubac, T., Bautista, J. E., Busca, N. G., Rich, J., Kirkby, D., Bailey, S., Font-Ribera, A., Slosar, A., Lee, K.-G., Pieri, M. M., Hamilton, J.-C., Aubourg, É., Blomqvist, M., Bovy, J., Brinkmann, J., Carithers, W., Dawson, K. S., Eisenstein, D. J., Gontcho, S. G. A., Kneib, J.-P., Le Goff, J.-M., Margala, D., Miralda-Escudé, J., Myers, A. D., Nichol, R. C., Noterdaeme, P., O'Connell, R., Olmstead, M. D., Palanque-Delabrouille, N., Pâris, I., Petitjean, P., Ross, N. P., Rossi, G., Schlegel, D. J., Schneider, D. P., Weinberg, D. H., Yèche, C., & York, D. G. 2015, *A&A*, 574, A59
- DESI Collaboration, Aghamousa, A., Aguilar, J., Ahlen, S., Alam, S., Allen, L. E., Allende Prieto, C., Annis, J., Bailey, S., Balland, C., & et al. 2016, *ArXiv e-prints*
- Dijkstra, M., Haiman, Z., & Loeb, A. 2004, *ApJ*, 613, 646
- Dodelson, S. 2003, *Modern Cosmology*, ed. A. Press (Academic Press)
- Dodelson, S., & Widrow, L. M. 1993, *Sterile neutrinos as dark matter*, Tech. rep.
- Dolag, K., Borgani, S., Murante, G., & Springel, V. 2009, *MNRAS*, 399, 497
- Dombrowski, J., Fabis, F., & Bartelmann, M. 2017, *ArXiv e-prints*
- Draine, B. T. 2011, *Physics of the Interstellar and Intergalactic Medium*

- D'Souza, R., Vegetti, S., & Kauffmann, G. 2015, *MNRAS*, 454, 4027
- Dubois, Y., Pichon, C., Welker, C., Le Borgne, D., Devriendt, J., Laigle, C., Codis, S., Pogosyan, D., Arnouts, S., Benabed, K., Bertin, E., Blaizot, J., Bouchet, F., Cardoso, J.-F., Colombi, S., de Lapparent, V., Desjacques, V., Gavazzi, R., Kassin, S., Kimm, T., McCracken, H., Milliard, B., Peirani, S., Prunet, S., Rouberol, S., Silk, J., Slyz, A., Sousbie, T., Teyssier, R., Tresse, L., Treyer, M., Vibert, D., & Volonteri, M. 2014, *MNRAS*, 444, 1453
- Efstathiou, G., Bernstein, G., Tyson, J. A., Katz, N., & Guhathakurta, P. 1991, *ApJ*, 380, L47
- Efstathiou, G., Frenk, C. S., White, S. D. M., & Davis, M. 1988, *MNRAS*, 235, 715
- Ehlers, J., & Buchert, T. 1997, *General Relativity and Gravitation*, 29, 733
- Einasto, J. 1993, in *European Astronomical Society. 2-nd General Meeting, Extragalactic Astronomy and observational Cosmology. Under the High Patronage of Mr. Lech Waleca, The President of Poland. The Nicolaus Copernicus University, Torun, Poland, 18-21 August 1993*, p.26, 26
- Einstein, A. 1915a, *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin)*, Seite 844-847.
- . 1915b, *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin)*, Seite 778-786.
- . 1915c, *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin)*, Seite 799-801.
- . 1916, *Annalen der Physik*, 354, 769
- . 1917, *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin)*, Seite 142-152.
- Eisenstein, D. J., Weinberg, D. H., Agol, E., Aihara, H., Allende Prieto, C., Anderson, S. F., Arns, J. A., Aubourg, É., Bailey, S., Balbinot, E., & et al. 2011, *AJ*, 142, 72
- Fabian, A. C. 1999, *MNRAS*, 308, L39
- Fabis, F., Berg, D., Kozlikin, E., & Bartelmann, M. 2014, *ArXiv e-prints*
- Fan, X., Strauss, M. A., Becker, R. H., White, R. L., Gunn, J. E., Knapp, G. R., Richards, G. T., Schneider, D. P., Brinkmann, J., & Fukugita, M. 2006, *AJ*, 132, 117
- Faucher-Giguère, C.-A., Feldmann, R., Quataert, E., Kereš, D., Hopkins, P. F., & Murray, N. 2016, *MNRAS*, 461, L32
- Faucher-Giguère, C.-A., Hopkins, P. F., Kereš, D., Muratov, A. L., Quataert, E., & Murray, N. 2015, *MNRAS*, 449, 987
- Faucher-Giguère, C.-A., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008a, *ApJ*, 688, 85
- Faucher-Giguère, C.-A., Lidz, A., Zaldarriaga, M., & Hernquist, L. 2009a, *ApJ*, 703, 1416
- . 2009b, *ApJ*, 703, 1416

- Faucher-Giguère, C.-A., Prochaska, J. X., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008b, *ApJ*, 681, 831
- Fedeli, C., Moscardini, L., & Bartelmann, M. 2009, *A&A*, 500, 667
- Feynman, R. P. 1950, *Phys. Rev.*, 80, 440
- Fisher, K. B. 1993, in *Cosmic Velocity Fields*, ed. F. Bouchet & M. Lachieze-Rey, 177
- Font-Ribera, A., Arnau, E., Miralda-Escudé, J., Rollinde, E., Brinkmann, J., Brownstein, J. R., Lee, K.-G., Myers, A. D., Palanque-Delabrouille, N., Pâris, I., Petitjean, P., Rich, J., Ross, N. P., Schneider, D. P., & White, M. 2013, *JCAP*, 5, 18
- Font-Ribera, A., Kirkby, D., Busca, N., Miralda-Escudé, J., Ross, N. P., Slosar, A., Rich, J., Aubourg, É., Bailey, S., Bhardwaj, V., Bautista, J., Beutler, F., Bizyaev, D., Blomqvist, M., Brewington, H., Brinkmann, J., Brownstein, J. R., Carithers, B., Dawson, K. S., Delubac, T., Ebelke, G., Eisenstein, D. J., Ge, J., Kinemuchi, K., Lee, K.-G., Malanushenko, V., Malanushenko, E., Marchante, M., Margala, D., Muna, D., Myers, A. D., Noterdaeme, P., Oravetz, D., Palanque-Delabrouille, N., Pâris, I., Petitjean, P., Pieri, M. M., Rossi, G., Schneider, D. P., Simmons, A., Viel, M., Yeche, C., & York, D. G. 2014a, *JCAP*, 5, 027
- Font-Ribera, A., McDonald, P., & Miralda-Escudé, J. 2012a, *JCAP*, 1, 001
- Font-Ribera, A., McDonald, P., Mostek, N., Reid, B. A., Seo, H.-J., & Slosar, A. 2014b, *JCAP*, 5, 023
- Font-Ribera, A., & Miralda-Escudé, J. 2012, *JCAP*, 7, 28
- Font-Ribera, A., Miralda-Escudé, J., Arnau, E., Carithers, B., Lee, K.-G., Noterdaeme, P., Pâris, I., Petitjean, P., Rich, J., Rollinde, E., Ross, N. P., Schneider, D. P., White, M., & York, D. G. 2012b, *JCAP*, 11, 59
- Forster, D., Nelson, D. R., & Stephen, M. J. 1977, *Phys. Rev. A*, 16, 732
- Frieman, J. A., Bassett, B., Becker, A., Choi, C., Cinabro, D., DeJongh, F., Depoy, D. L., Dilday, B., Doi, M., Garnavich, P. M., Hogan, C. J., Holtzman, J., Im, M., Jha, S., Kessler, R., Konishi, K., Lampeitl, H., Marriner, J., Marshall, J. L., McGinnis, D., Miknaitis, G., Nichol, R. C., Prieto, J. L., Riess, A. G., Richmond, M. W., Romani, R., Sako, M., Schneider, D. P., Smith, M., Takanashi, N., Tokita, K., van der Heyden, K., Yasuda, N., Zheng, C., Adelman-McCarthy, J., Annis, J., Assef, R. J., Barentine, J., Bender, R., Blandford, R. D., Boroski, W. N., Bremer, M., Brewington, H., Collins, C. A., Crotts, A., Dembicky, J., Eastman, J., Edge, A., Edmondson, E., Elson, E., Eyler, M. E., Filippenko, A. V., Foley, R. J., Frank, S., Goobar, A., Gueth, T., Gunn, J. E., Harvanek, M., Hopp, U., Ihara, Y., Ivezić, Ž., Kahn, S., Kaplan, J., Kent, S., Ketzeback, W., Kleinman, S. J., Kollatschny, W., Kron, R. G., Krzesiński, J., Lamenti, D., Leloudas, G., Lin, H., Long, D. C., Lucey, J., Lupton, R. H., Malanushenko, E., Malanushenko, V., McMillan, R. J., Mendez, J., Morgan, C. W., Morokuma, T., Nitta, A., Ostman, L., Pan, K., Rockosi, C. M., Romer, A. K., Ruiz-Lapuente, P., Saurage, G., Schlesinger, K., Snedden, S. A., Sollerman, J., Stoughton, C., Stritzinger, M., Subba Rao, M., Tucker, D., Vaisanen, P., Watson, L. C., Watters, S., Wheeler, J. C., Yanny, B., & York, D. 2008, *AJ*, 135, 338
- Fukugita, M., Hogan, C. J., & Peebles, P. J. E. 1998, *ApJ*, 503, 518

- Fumagalli, M., Hennawi, J. F., Prochaska, J. X., Kasen, D., Dekel, A., Ceverino, D., & Primack, J. 2014, *ApJ*, 780, 74
- Furlanetto, S. R., & Oh, S. P. 2008, *ApJ*, 681, 1
- Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, *Phys. Rep.*, 433, 181
- Gaikwad, P., Srianand, R., Choudhury, T. R., & Khaire, V. 2017, *MNRAS*, 467, 3172
- Gao, L., Yoshida, N., Abel, T., Frenk, C. S., Jenkins, A., & Springel, V. 2007, *MNRAS*, 378, 449
- Gardner, J. P., Mather, J. C., Clampin, M., Doyon, R., Greenhouse, M. A., Hammel, H. B., Hutchings, J. B., Jakobsen, P., Lilly, S. J., Long, K. S., Lunine, J. I., McCaughrean, M. J., Mountain, M., Nella, J., Rieke, G. H., Rieke, M. J., Rix, H.-W., Smith, E. P., Sonneborn, G., Stiavelli, M., Stockman, H. S., Windhorst, R. A., & Wright, G. S. 2006, in *Proc. SPIE*, Vol. 6265, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 62650N
- Genel, S., Vogelsberger, M., Springel, V., Sijacki, D., Nelson, D., Snyder, G., Rodriguez-Gomez, V., Torrey, P., & Hernquist, L. 2014, *MNRAS*, 445, 175
- Giallongo, E., Grazian, A., Fiore, F., Fontana, A., Pentericci, L., Vanzella, E., Dickinson, M., Kocevski, D., Castellano, M., Cristiani, S., Ferguson, H., Finkelstein, S., Grogin, N., Hathi, N., Koekemoer, A. M., Newman, J. A., & Salvato, M. 2015, *A&A*, 578, A83
- Giodini, S., Pierini, D., Finoguenov, A., Pratt, G. W., Boehringer, H., Leauthaud, A., Guzzo, L., Aussel, H., Bolzonella, M., Capak, P., Elvis, M., Hasinger, G., Ilbert, O., Kartaltepe, J. S., Koekemoer, A. M., Lilly, S. J., Massey, R., McCracken, H. J., Rhodes, J., Salvato, M., Sanders, D. B., Scoville, N. Z., Sasaki, S., Smolcic, V., Taniguchi, Y., Thompson, D., & COSMOS Collaboration. 2009, *ApJ*, 703, 982
- Gnedin, N. Y., & Hollon, N. 2012, *ApJS*, 202, 13
- Gnedin, N. Y., & Hui, L. 1996, *ApJ*, 472, L73
- . 1998, *MNRAS*, 296, 44
- Gontcho A Gontcho, S., Miralda-Escudé, J., & Busca, N. G. 2014, *MNRAS*, 442, 187
- Guha Sarkar, T., & Datta, K. K. 2015, *JCAP*, 8, 1
- Gunn, J. E., & Peterson, B. A. 1965, *ApJ*, 142, 1633
- Guo, Q., White, S., Boylan-Kolchin, M., De Lucia, G., Kauffmann, G., Lemson, G., Li, C., Springel, V., & Weinmann, S. 2011, *MNRAS*, 413, 101
- Gutcke, T. A., Stinson, G. S., Macciò, A. V., Wang, L., & Dutton, A. A. 2017, *MNRAS*, 464, 2796
- Guth, A. H. 1981, *Phys. Rev. D*, 23, 347
- Haardt, F., & Madau, P. 2012, *ApJ*, 746, 125
- Habib, S., Morozov, V., Finkel, H., Pope, A., Heitmann, K., Kumaran, K., Peterka, T., Insley, J., Daniel, D., Fasel, P., Frontiere, N., & Lukic, Z. 2012, *ArXiv e-prints*

- Habib, S., Morozov, V., Frontiere, N., Finkel, H., Pope, A., & Heitmann, K. 2013, in SC '13 Proceedings of SC13: International Conference for High Performance Computing, Networking, Storage and Analysis, Article #6
- Hahn, O., & Abel, T. 2011, *MNRAS*, 415, 2101
- Hamana, T., Colombi, S., & Suto, Y. 2001, *A&A*, 367, 18
- Hambly, N., Read, M., Mann, R., Sutorius, E., Bond, I., MacGillivray, H., Williams, P., & Lawrence, A. 2004, in *Astronomical Society of the Pacific Conference Series*, Vol. 314, *Astronomical Data Analysis Software and Systems (ADASS) XIII*, ed. F. Ochsenbein, M. G. Allen, & D. Egret, 137
- Hambly, N. C., Davenhall, A. C., Irwin, M. J., & MacGillivray, H. T. 2001a, *MNRAS*, 326, 1315
- Hambly, N. C., Irwin, M. J., & MacGillivray, H. T. 2001b, *MNRAS*, 326, 1295
- Hambly, N. C., MacGillivray, H. T., Read, M. A., Tritton, S. B., Thomson, E. B., Kelly, B. D., Morgan, D. H., Smith, R. E., Driver, S. P., Williamson, J., Parker, Q. A., Hawkins, M. R. S., Williams, P. M., & Lawrence, A. 2001c, *MNRAS*, 326, 1279
- Hamilton, A. J. S., Kumar, P., Lu, E., & Matthews, A. 1991, *ApJ*, 374, L1
- Hamilton, A. J. S., & Tegmark, M. 2000, *MNRAS*, 312, 285
- Hazard, C., Mackey, M. B., & Shimmins, A. J. 1963, *Nature*, 197, 1037
- Heckman, T. M., & Thompson, T. A. 2017, *ArXiv e-prints*
- Heitmann, K., Higdon, D., White, M., Habib, S., Williams, B. J., Lawrence, E., & Wagner, C. 2009, *ApJ*, 705, 156
- Heitmann, K., Lawrence, E., Kwan, J., Habib, S., & Higdon, D. 2014, *ApJ*, 780, 111
- Heitmann, K., White, M., Wagner, C., Habib, S., & Higdon, D. 2010, *ApJ*, 715, 104
- Hennawi, J. F. 2004, PhD thesis, Ph.D dissertation, 2004. 232 pages; United States – New Jersey: Princeton University; 2004. Publication Number: AAT 3151085. DAI-B 65/10, p. 5189, Apr 2005
- Hennawi, J. F., Myers, A. D., Shen, Y., Strauss, M. A., Djorgovski, S. G., Fan, X., Glikman, E., Mahabal, A., Martin, C. L., Richards, G. T., Schneider, D. P., & Shankar, F. 2010, *ApJ*, 719, 1672
- Hennawi, J. F., & Prochaska, J. X. 2007, *ApJ*, 655, 735
- Hennawi, J. F., Prochaska, J. X., Burles, S., Strauss, M. A., Richards, G. T., Schlegel, D. J., Fan, X., Schneider, D. P., Zakamska, N. L., Oguri, M., Gunn, J. E., Lupton, R. H., & Brinkmann, J. 2006a, *ApJ*, 651, 61
- Hennawi, J. F., Strauss, M. A., Oguri, M., Inada, N., Richards, G. T., Pindor, B., Schneider, D. P., Becker, R. H., Gregg, M. D., Hall, P. B., Johnston, D. E., Fan, X., Burles, S., Schlegel, D. J., Gunn, J. E., Lupton, R. H., Bahcall, N. A., Brunner, R. J., & Brinkmann, J. 2006b, *AJ*, 131, 1
- Hernquist, L., Katz, N., Weinberg, D. H., & Miralda-Escudé, J. 1996, *ApJ*, 457, L51

- Hinshaw, G., Larson, D., Komatsu, E., Spergel, D. N., Bennett, C. L., Dunkley, J., Nolte, M. R., Halpern, M., Hill, R. S., Odegard, N., Page, L., Smith, K. M., Weiland, J. L., Gold, B., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., Wollack, E., & Wright, E. L. 2013, *ApJS*, 208, 19
- Hirata, C. M. 2017, *ArXiv e-prints*
- Hockney, R. W., & Eastwood, J. W. 1981, *Computer Simulation Using Particles*
- Hopkins, P. F., Kereš, D., Oñorbe, J., Faucher-Giguère, C.-A., Quataert, E., Murray, N., & Bullock, J. S. 2014, *MNRAS*, 445, 581
- Howlett, C., Lewis, A., Hall, A., & Challinor, A. 2012, *JCAP*, 4, 027
- Howlett, C., Staveley-Smith, L., Elahi, P. J., Hong, T., Jarrett, T. H., Jones, D. H., Koribalski, B. S., Macri, L. M., Masters, K. L., & Springob, C. M. 2017, *ArXiv e-prints*
- Hubble, E. P. 1925, *ApJ*, 62
- Hui, L., & Gnedin, N. Y. 1997, *MNRAS*, 292, 27
- Hwang, J.-c., Jeong, D., & Noh, H. 2016, *MNRAS*, 459, 1104
- Iliev, I. T., Pen, U.-L., Bond, J. R., Mellema, G., & Shapiro, P. R. 2007, *ApJ*, 660, 933
- Jain, B. 1997, *MNRAS*, 287, 687
- Jain, B., & Bertschinger, E. 1998, *ApJ*, 509, 517
- Jenkins, A., Frenk, C. S., Pearce, F. R., Thomas, P. A., Colberg, J. M., White, S. D. M., Couchman, H. M. P., Peacock, J. A., Efstathiou, G., & Nelson, A. H. 1998, *ApJ*, 499, 20
- Jennings, E., Baugh, C. M., Li, B., Zhao, G.-B., & Koyama, K. 2012, *MNRAS*, 425, 2128
- Jennings, E., Baugh, C. M., & Pascoli, S. 2011, *MNRAS*, 410, 2081
- Johnson, A., Blake, C., Koda, J., Ma, Y.-Z., Colless, M., Crocce, M., Davis, T. M., Jones, H., Magoulas, C., Lucey, J. R., Mould, J., Scrimgeour, M. I., & Springob, C. M. 2014, *MNRAS*, 444, 3926
- Johnson, J. L., & Khochfar, S. 2011, *ApJ*, 743, 126
- Jones, D. H., Read, M. A., Saunders, W., Colless, M., Jarrett, T., Parker, Q. A., Fairall, A. P., Mauch, T., Sadler, E. M., Watson, F. G., Burton, D., Campbell, L. A., Cass, P., Croom, S. M., Dawe, J., Fiegert, K., Frankcombe, L., Hartley, M., Huchra, J., James, D., Kirby, E., Lahav, O., Lucey, J., Mamon, G. A., Moore, L., Peterson, B. A., Prior, S., Proust, D., Russell, K., Safouris, V., Wakamatsu, K.-I., Westra, E., & Williams, M. 2009, *MNRAS*, 399, 683
- Jones, D. H., Saunders, W., Colless, M., Read, M. A., Parker, Q. A., Watson, F. G., Campbell, L. A., Burkey, D., Mauch, T., Moore, L., Hartley, M., Cass, P., James, D., Russell, K., Fiegert, K., Dawe, J., Huchra, J., Jarrett, T., Lahav, O., Lucey, J., Mamon, G. A., Proust, D., Sadler, E. M., & Wakamatsu, K.-i. 2004, *MNRAS*, 355, 747

- Kaiser, N., & Peacock, J. A. 1991, *ApJ*, 379, 482
- Kannan, R., Stinson, G. S., Macciò, A. V., Hennawi, J. F., Woods, R., Wadsley, J., Shen, S., Robitaille, T., Cantalupo, S., Quinn, T. R., & Christensen, C. 2014, *MNRAS*, 437, 2882
- Kaurov, A. A., & Gnedin, N. Y. 2014, *ApJ*, 787, 146
- Kembhavi, A. K., & Narlikar, J. V. 1999, *Quasars and active galactic nuclei : an introduction*
- Khurshudyan, M., & Khurshudyan, A. 2017, *ArXiv e-prints*
- Kim, T.-S., Carswell, R. F., Cristiani, S., D'Odorico, S., & Giallongo, E. 2002, *MNRAS*, 335, 555
- King, A. 2003, *ApJ*, 596, L27
- Kollmeier, J. A., Miralda-Escudé, J., Cen, R., & Ostriker, J. P. 2006, *ApJ*, 638, 52
- Kollmeier, J. A., Weinberg, D. H., Davé, R., & Katz, N. 2003, *ApJ*, 594, 75
- Komatsu, E., Smith, K. M., Dunkley, J., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Nolte, M. R., Page, L., Spergel, D. N., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Odegard, N., Tucker, G. S., Weiland, J. L., Wollack, E., & Wright, E. L. 2011, *ApJS*, 192, 18
- Kormendy, J., & Ho, L. C. 2013, *ARA&A*, 51, 511
- Kozlikin, E., Fabis, F., Lilow, R., Viermann, C., & Bartelmann, M. 2014, *ArXiv e-prints*
- Kuhlen, M., & Faucher-Giguère, C.-A. 2012, *MNRAS*, 423, 862
- Kulkarni, G., Hennawi, J. F., Oñorbe, J., Rorai, A., & Springel, V. 2015, *ArXiv e-prints*
- Lau, M. W., Prochaska, J. X., & Hennawi, J. F. 2016, *ApJS*, 226, 25
- Laureijs, R., Amiaux, J., Arduini, S., Auguères, J. ., Brinchmann, J., Cole, R., Cropper, M., Dabin, C., Duvet, L., Ealet, A., & et al. 2011, *ArXiv e-prints*
- Lawrence, E., Heitmann, K., White, M., Higdon, D., Wagner, C., Habib, S., & Williams, B. 2010, *ApJ*, 713, 1322
- Lee, K.-G. 2012, *ApJ*, 753, 136
- Lee, K.-G., Hennawi, J. F., Spergel, D. N., Weinberg, D. H., Hogg, D. W., Viel, M., Bolton, J. S., Bailey, S., Pieri, M. M., Carithers, W., Schlegel, D. J., Lundgren, B., Palanque-Delabrouille, N., Suzuki, N., Schneider, D. P., & Yèche, C. 2015, *ApJ*, 799, 196
- Lee, K.-G., Hennawi, J. F., Stark, C., Prochaska, J. X., White, M., Schlegel, D. J., Eilers, A.-C., Arinyo-i-Prats, A., Suzuki, N., Croft, R. A. C., Caputi, K. I., Cassata, P., Ilbert, O., Garilli, B., Koekemoer, A. M., Le Brun, V., Le Fèvre, O., Maccagni, D., Nugent, P., Taniguchi, Y., Tasca, L. A. M., Tresse, L., Zamorani, G., & Zucca, E. 2014, *ApJ*, 795, L12
- Lewis, A., Challinor, A., & Lasenby, A. 2000, *ApJ*, 538, 473

- Li, B., Hellwing, W. A., Koyama, K., Zhao, G.-B., Jennings, E., & Baugh, C. M. 2013, *MNRAS*, 428, 743
- Lidz, A., Faucher-Giguère, C.-A., Dall’Aglio, A., McQuinn, M., Fechner, C., Zaldarriaga, M., Hernquist, L., & Dutta, S. 2010, *ApJ*, 718, 199
- Lilow et al. in prep.
- Littek, Sorini et al. in prep.
- Lochhaas, C., Weinberg, D. H., Peirani, S., Dubois, Y., Colombi, S., Blaizot, J., Font-Ribera, A., Pichon, C., & Devriendt, J. 2015, *ArXiv e-prints*
- Lovisari, L., Reiprich, T. H., & Schellenberger, G. 2015, *A&A*, 573, A118
- Lukić, Z., Stark, C. W., Nugent, P., White, M., Meiksin, A. A., & Almgren, A. 2015, *MNRAS*, 446, 3697
- Lundgren, B., Kinemuchi, K., Zasowski, G., Lucatello, S., Diamond-Stanic, A. M., Tremonti, C. A., Myers, A. D., Aragón-Salamanca, A., Gillespie, B., Ho, S., & Gallagher, J. S. 2015, *PASP*, 127, 776
- Lynds, R. 1971, *ApJ*, 164, L73
- Ma, C.-P., & Bertschinger, E. 1995, *ApJ*, 455, 7
- Ma, C.-P., & Fry, J. N. 2000, *ApJ*, 543, 503
- Madau, P., & Dickinson, M. 2014, *ARA&A*, 52, 415
- Madau, P., & Haardt, F. 2015, *ApJ*, 813, L8
- Madau, P., Haardt, F., & Rees, M. J. 1999, *ApJ*, 514, 648
- Martin, P. C., Siggia, E. D., & Rose, H. A. 1973, *Phys. Rev. A*, 8, 423
- Matarrese, S., Coles, P., Lucchin, F., & Moscardini, L. 1997, *MNRAS*, 286, 115
- Mazenko, G. F. 2010, *Phys. Rev. E*, 81, 061102
- . 2011, *Phys. Rev. E*, 83, 041125
- McConnell, N. J., & Ma, C.-P. 2013, *ApJ*, 764, 184
- McCourt, M., Oh, S. P., O’Leary, R. M., & Madigan, A.-M. 2016, *ArXiv e-prints*
- McDonald, P. 2003, *ApJ*, 585, 34
- McDonald, P., Miralda-Escudé, J., Rauch, M., Sargent, W. L. W., Barlow, T. A., Cen, R., & Ostriker, J. P. 2000, *ApJ*, 543, 1
- McDonald, P., Seljak, U., Burles, S., Schlegel, D. J., Weinberg, D. H., Cen, R., Shih, D., Schaye, J., Schneider, D. P., Bahcall, N. A., Briggs, J. W., Brinkmann, J., Brunner, R. J., Fukugita, M., Gunn, J. E., Ivezić, Ž., Kent, S., Lupton, R. H., & Vanden Berk, D. E. 2006, *ApJS*, 163, 80
- McDonald, P., Seljak, U., Cen, R., Shih, D., Weinberg, D. H., Burles, S., Schneider, D. P., Schlegel, D. J., Bahcall, N. A., Briggs, J. W., Brinkmann, J., Fukugita, M., Ivezić, Ž., Kent, S., & Vanden Berk, D. E. 2005, *ApJ*, 635, 761

- McGreer, I. D., Mesinger, A., & D'Odorico, V. 2015, MNRAS, 447, 499
- McQuinn, M. 2012, MNRAS, 426, 1349
- . 2016, ARA&A, 54, 313
- McQuinn, M., Hernquist, L., Lidz, A., & Zaldarriaga, M. 2011, MNRAS, 415, 977
- McQuinn, M., & White, M. 2011, MNRAS, 415, 2257
- Meiksin, A., Bolton, J. S., & Puchwein, E. 2017, MNRAS
- Meiksin, A., Bolton, J. S., & Tittley, E. R. 2014, MNRAS, 445, 2462
- . 2015, MNRAS, 453, 899
- Meiksin, A., & White, M. 1999, MNRAS, 308, 1179
- . 2001, MNRAS, 324, 141
- Meiksin, A. A. 2009, Reviews of Modern Physics, 81, 1405
- Mirabel, I. F., Dijkstra, M., Laurent, P., Loeb, A., & Pritchard, J. R. 2011, A&A, 528, A149
- Miralda-Escudé, J. 2003, ApJ, 597, 66
- Moscardini, L., Matarrese, S., Andreani, P., Bartelmann, M., & Mo, H. 2002, in *Astronomical Society of the Pacific Conference Series*, Vol. 268, *Tracing Cosmic Evolution with Galaxy Clusters*, ed. S. Borgani, M. Mezzetti, & R. Valdarnini, 86
- Moster, B. P., Naab, T., & White, S. D. M. 2013, MNRAS, 428, 3121
- Muñoz, J. A., & Loeb, A. 2008, MNRAS, 386, 2323
- Naoz, S., Noter, S., & Barkana, R. 2006, MNRAS, 373, L98
- Nelson, D., Genel, S., Pillepich, A., Vogelsberger, M., Springel, V., & Hernquist, L. 2016, MNRAS, 460, 2881
- Nelson, D., Pillepich, A., Genel, S., Vogelsberger, M., Springel, V., Torrey, P., Rodriguez-Gomez, V., Sijacki, D., Snyder, G. F., Griffen, B., Marinacci, F., Blecha, L., Sales, L., Xu, D., & Hernquist, L. 2015, *Astronomy and Computing*, 13, 12
- Neuschaefter, L. W., Windhorst, R. A., & Dressler, A. 1991, ApJ, 382, 32
- Neyman, J., & Scott, E. L. 1952, ApJ, 116, 144
- Nishioka, H., & Yamamoto, K. 1999, ApJ, 520, 426
- . 2000, ApJS, 128, 1
- Norman, M. L., Paschos, P., & Harkness, R. 2009, *Journal of Physics Conference Series*, 180, 012021
- Noterdaeme, P., Petitjean, P., Carithers, W. C., Pâris, I., Font-Ribera, A., Bailey, S., Aubourg, E., Bizyaev, D., Ebelke, G., Finley, H., Ge, J., Malanushenko, E., Malanushenko, V., Miralda-Escudé, J., Myers, A. D., Oravetz, D., Pan, K., Pieri, M. M., Ross, N. P., Schneider, D. P., Simmons, A., & York, D. G. 2012, A&A, 547, L1

- Oñorbe, J., Garrison-Kimmel, S., Maller, A. H., Bullock, J. S., Rocha, M., & Hahn, O. 2014, *MNRAS*, 437, 1894
- Occhionero, F., Achilli, S., Scaramella, R., & Vittorio, N. 1984, in *BAAS*, Vol. 16, *Bulletin of the American Astronomical Society*, 487
- Oesch, P. A., Bouwens, R. J., Illingworth, G. D., Franx, M., Ammons, S. M., van Dokkum, P. G., Trenti, M., & Labbé, I. 2015, *ApJ*, 808, 104
- Oh, S. P. 2001, *ApJ*, 553, 499
- Oort, J. H. 1932, *Bull. Astron. Inst. Netherlands*, 6, 249
- Oppenheimer, B. D., & Davé, R. 2009, *MNRAS*, 395, 1875
- Padmanabhan, T. 2006, in *American Institute of Physics Conference Series*, Vol. 843, *Graduate School in Astronomy: X*, ed. S. Daffon, J. Alcaniz, E. Telles, & R. de la Reza, 111–166
- Padmanabhan, T., Cen, R., Ostriker, J. P., & Summers, F. J. 1996, *ApJ*, 466, 604
- Palanque-Delabrouille, N., Yèche, C., Baur, J., Magneville, C., Rossi, G., Lesgourgues, J., Borde, A., Burtin, E., LeGoff, J.-M., Rich, J., Viel, M., & Weinberg, D. 2015, *JCAP*, 11, 011
- Park, C., Gott, III, J. R., & da Costa, L. N. 1992, *ApJ*, 392, L51
- Parker, Q. A., Phillipps, S., Pierce, M. J., Hartley, M., Hambly, N. C., Read, M. A., MacGillivray, H. T., Tritton, S. B., Cass, C. P., Cannon, R. D., Cohen, M., Drew, J. E., Frew, D. J., Hopewell, E., Mader, S., Malin, D. F., Masheder, M. R. W., Morgan, D. H., Morris, R. A. H., Russeil, D., Russell, K. S., & Walker, R. N. F. 2005, *MNRAS*, 362, 689
- Peacock, J. A., & Dodds, S. J. 1994, *MNRAS*, 267, 1020
- . 1996, *MNRAS*, 280, L19
- Peacock, J. A., & Nicholson, D. 1991, *MNRAS*, 253, 307
- Peacock, J. A., & Smith, R. E. 2000, *MNRAS*, 318, 1144
- Peacock, J. A., & West, M. J. 1992, *MNRAS*, 259, 494
- Peebles, P. J. E. 1973, *ApJ*, 185, 413
- . 1980, *The large-scale structure of the universe*
- Peebles, P. J. E., & Hauser, M. G. 1974, *ApJS*, 28, 19
- Peirani, S., Weinberg, D. H., Colombi, S., Blaizot, J., Dubois, Y., & Pichon, C. 2014, *ApJ*, 784, 11
- Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., Castro, P. G., Deustua, S., Fabbro, S., Goobar, A., Groom, D. E., Hook, I. M., Kim, A. G., Kim, M. Y., Lee, J. C., Nunes, N. J., Pain, R., Pennypacker, C. R., Quimby, R., Lidman, C., Ellis, R. S., Irwin, M., McMahon, R. G., Ruiz-Lapuente, P., Walton, N., Schaefer, B., Boyle, B. J., Filippenko, A. V., Matheson, T., Fruchter, A. S., Panagia, N., Newberg, H. J. M., Couch, W. J., & Project, T. S. C. 1999, *ApJ*, 517, 565

- Péroux, C., Dessauges-Zavadsky, M., D'Odorico, S., Sun Kim, T., & McMahon, R. G. 2005, *MNRAS*, 363, 479
- Peter, A. H. G. 2012, ArXiv e-prints
- Petitjean, P., Mueket, J. P., & Kates, R. E. 1995, *A&A*, 295, L9
- Petkova, M., & Springel, V. 2011, *MNRAS*, 412, 935
- Pillepich, A., Springel, V., Nelson, D., Genel, S., Naiman, J., Pakmor, R., Hernquist, L., Torrey, P., Vogelsberger, M., Weinberger, R., & Marinacci, F. 2017, ArXiv e-prints
- Planck Collaboration, Ade, P. A. R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A. J., Barreiro, R. B., Bartlett, J. G., & et al. 2016, *A&A*, 594, A13
- Pontzen, A. 2014, *Phys. Rev. D*, 89, 083010
- Primack, J. R. 1984, in *Particles and Gravity*, ed. G. Domokos & S. Kovesi-Domokos, 175–190
- Prochaska, J. X., & Hennawi, J. F. 2009, *ApJ*, 690, 1558
- Prochaska, J. X., Hennawi, J. F., Lee, K.-G., Cantalupo, S., Bovy, J., Djorgovski, S. G., Ellison, S. L., Lau, M. W., Martin, C. L., Myers, A., Rubin, K. H. R., & Simcoe, R. A. 2013, *ApJ*, 776, 136
- Prochaska, J. X., Lau, M. W., & Hennawi, J. F. 2014, *ApJ*, 796, 140
- Prochaska, J. X., Weiner, B., Chen, H.-W., Mulchaey, J., & Cooksey, K. 2011, *ApJ*, 740, 91
- Prochaska, J. X., & Wolfe, A. M. 2009, *ApJ*, 696, 1543
- Raccanelli, A., Bertacca, D., Doré, O., & Maartens, R. 2014, *JCAP*, 8, 22
- Rahmati, A., Pawlik, A. H., Raičević, M., & Schaye, J. 2013, *MNRAS*, 430, 2427
- Rahmati, A., Schaye, J., Bower, R. G., Crain, R. A., Furlong, M., Schaller, M., & Theuns, T. 2015, *MNRAS*, 452, 2034
- Rahmati, A., Schaye, J., Crain, R. A., Oppenheimer, B. D., Schaller, M., & Theuns, T. 2016, *MNRAS*, 459, 310
- Raichoor, A., Comparat, J., Delubac, T., Kneib, J.-P., Yèche, C., Dawson, K. S., Percival, W. J., Dey, A., Lang, D., Schlegel, D. J., Gorgoni, C., Bautista, J., Brownstein, J. R., Mariappan, V., Seo, H.-J., Tinker, J. L., Ross, A. J., Wang, Y., Zhao, G.-B., Moustakas, J., Palanque-Delabrouille, N., Jullo, E., Newmann, J. A., Prada, F., & Zhu, G. B. 2017, ArXiv e-prints
- Rakic, O., Schaye, J., Steidel, C. C., Booth, C. M., Dalla Vecchia, C., & Rudie, G. C. 2013, *MNRAS*, 433, 3103
- Rakic, O., Schaye, J., Steidel, C. C., & Rudie, G. C. 2012, *ApJ*, 751, 94
- Rauch, M. 1998, *ARA&A*, 36, 267

- Rauch, M., Miralda-Escudé, J., Sargent, W. L. W., Barlow, T. A., Weinberg, D. H., Hernquist, L., Katz, N., Cen, R., & Ostriker, J. P. 1997, *ApJ*, 489, 7
- Rees, M. J., & Ostriker, J. P. 1977, *MNRAS*, 179, 541
- Riess, A. G., & *et al.*, B. P. S. 1998, *The Astronomical Journal*
- Rorai, A., Hennawi, J. F., Oñorbe, J., White, M., Prochaska, J. X., Kulkarni, G., Walther, M., Lukić, Z., & Lee, K.-G. 2017, *Science*, 356, 418
- Rorai, A., Hennawi, J. F., & White, M. 2013, *ApJ*, 775, 81
- Rubin, K. H. R., Hennawi, J. F., Prochaska, J. X., Simcoe, R. A., Myers, A., & Lau, M. W. 2015, *ApJ*, 808, 38
- Rubin, V. C., Ford, Jr., W. K., & Thonnard, N. 1980, *ApJ*, 238, 471
- Rudie, G. C., Steidel, C. C., & Pettini, M. 2012a, *ApJ*, 757, L30
- Rudie, G. C., Steidel, C. C., Shapley, A. E., & Pettini, M. 2013, *ApJ*, 769, 146
- Rudie, G. C., Steidel, C. C., Trainor, R. F., Rakic, O., Bogosavljević, M., Pettini, M., Reddy, N., Shapley, A. E., Erb, D. K., & Law, D. R. 2012b, *ApJ*, 750, 67
- Sargent, W. L. W., Young, P. J., Boksenberg, A., & Tytler, D. 1980, *ApJS*, 42, 41
- Schaye, J., Crain, R. A., Bower, R. G., Furlong, M., Schaller, M., Theuns, T., Dalla Vecchia, C., Frenk, C. S., McCarthy, I. G., Helly, J. C., Jenkins, A., Rosas-Guevara, Y. M., White, S. D. M., Baes, M., Booth, C. M., Camps, P., Navarro, J. F., Qu, Y., Rahmati, A., Sawala, T., Thomas, P. A., & Trayford, J. 2015, *MNRAS*, 446, 521
- Schaye, J., Dalla Vecchia, C., Booth, C. M., Wiersma, R. P. C., Theuns, T., Haas, M. R., Bertone, S., Duffy, A. R., McCarthy, I. G., & van de Voort, F. 2010, *MNRAS*, 402, 1536
- Scherrer, R. J., & Bertschinger, E. 1991, *ApJ*, 381, 349
- Schmidt, B. P., Suntzeff, N. B., Phillips, M. M., Schommer, R. A., Clocchiatti, A., Kirshner, R. P., Garnavich, P., Challis, P., Leibundgut, B., Spyromilio, J., Riess, A. G., Filippenko, A. V., Hamuy, M., Smith, R. C., Hogan, C., Stubbs, C., Diercks, A., Reiss, D., Gilliland, R., Tonry, J., Maza, J., Dressler, A., Walsh, J., & Ciardullo, R. 1998, *ApJ*, 507, 46
- Schmidt, M. 1963, *Nature*, 197, 1040
- Schneider, A., Teyssier, R., Potter, D., Stadel, J., Onions, J., Reed, D. S., Smith, R. E., Springel, V., Pearce, F. R., & Scoccimarro, R. 2016, *JCAP*, 4, 047
- Scoccimarro, R., Sheth, R. K., Hui, L., & Jain, B. 2001, *ApJ*, 546, 20
- Seljak, U. 2000, *MNRAS*, 318, 203
- Seljak, U., Makarov, A., McDonald, P., Anderson, S. F., Bahcall, N. A., Brinkmann, J., Burles, S., Cen, R., Doi, M., Gunn, J. E., Ivezić, Ž., Kent, S., Loveday, J., Lupton, R. H., Munn, J. A., Nichol, R. C., Ostriker, J. P., Schlegel, D. J., Schneider, D. P., Tegmark, M., Berk, D. E., Weinberg, D. H., & York, D. G. 2005, *Phys. Rev. D*, 71, 103515

- Seljak, U., & Zaldarriaga, M. 1996, *ApJ*, 469, 437
- Shapiro, P. R., Giroux, M. L., & Babul, A. 1994, *ApJ*, 427, 25
- Shen, S., Madau, P., Guedes, J., Mayer, L., Prochaska, J. X., & Wadsley, J. 2013, *ApJ*, 765, 89
- Shen, S., Mo, H. J., White, S. D. M., Blanton, M. R., Kauffmann, G., Voges, W., Brinkmann, J., & Csabai, I. 2003, *MNRAS*, 343, 978
- Shull, J. M., France, K., Danforth, C. W., Smith, B., & Tumlinson, J. 2010, *ApJ*, 722, 1312
- Sijacki, D., Springel, V., Di Matteo, T., & Hernquist, L. 2007, *MNRAS*, 380, 877
- Sijacki, D., Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Snyder, G. F., Nelson, D., & Hernquist, L. 2015, *MNRAS*, 452, 575
- Silk, J., & Rees, M. J. 1998, *A&A*, 331, L1
- Simcoe, R. A., Sargent, W. L. W., Rauch, M., & Becker, G. 2006, *ApJ*, 637, 648
- Skillman, S. W., Warren, M. S., Turk, M. J., Wechsler, R. H., Holz, D. E., & Sutter, P. M. 2014, *ArXiv e-prints*
- Skrutskie, M. F., Cutri, R. M., Stiening, R., Weinberg, M. D., Schneider, S., Carpenter, J. M., Beichman, C., Capps, R., Chester, T., Elias, J., Huchra, J., Liebert, J., Lonsdale, C., Monet, D. G., Price, S., Seitzer, P., Jarrett, T., Kirkpatrick, J. D., Gizis, J. E., Howard, E., Evans, T., Fowler, J., Fullmer, L., Hurt, R., Light, R., Kopan, E. L., Marsh, K. A., McCallon, H. L., Tam, R., Van Dyk, S., & Wheelock, S. 2006, *AJ*, 131, 1163
- Slosar, A., Font-Ribera, A., Pieri, M. M., Rich, J., Le Goff, J.-M., Aubourg, É., Brinkmann, J., Busca, N., Carithers, B., Charlassier, R., Cortès, M., Croft, R., Dawson, K. S., Eisenstein, D., Hamilton, J.-C., Ho, S., Lee, K.-G., Lupton, R., McDonald, P., Medolin, B., Muna, D., Miralda-Escudé, J., Myers, A. D., Nichol, R. C., Palanque-Delabrouille, N., Pâris, I., Petitjean, P., Piškur, Y., Rollinde, E., Ross, N. P., Schlegel, D. J., Schneider, D. P., Sheldon, E., Weaver, B. A., Weinberg, D. H., Yèche, C., & York, D. G. 2011, *JCAP*, 9, 001
- Slosar, A., Ho, S., White, M., & Louis, T. 2009, *JCAP*, 10, 19
- Smith, R. E., Peacock, J. A., Jenkins, A., White, S. D. M., Frenk, C. S., Pearce, F. R., Thomas, P. A., Efstathiou, G., & Couchman, H. M. P. 2003, *MNRAS*, 341, 1311
- Somerville, R. S., & Davé, R. 2015, *ARA&A*, 53, 51
- Songaila, A. 2006, *AJ*, 131, 24
- Sorini, D. 2017, *JCAP*, 4, 029
- Sorini, D., Oñorbe, J., Lukić, Z., & Hennawi, J. F. 2016, *ApJ*, 827, 97
- Sorini, D., Oñorbe, J., Hennawi, J. F., & Lukić, Z. 2017, to be submitted
- Sorini et al. in prep.

- Spergel, D. N., Verde, L., Peiris, H. V., Komatsu, E., Nolta, M. R., Bennett, C. L., Halpern, M., Hinshaw, G., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Page, L., Tucker, G. S., Weiland, J. L., Wollack, E., & Wright, E. L. 2003, *ApJS*, 148, 175
- Springel, V. 2000, *MNRAS*, 312, 859
- . 2005, *MNRAS*, 364, 1105
- . 2010a, *MNRAS*, 401, 791
- . 2010b, *ARA&A*, 48, 391
- . 2011, ArXiv e-prints
- . 2012, *Astronomische Nachrichten*, 333, 515
- Springel, V., Di Matteo, T., & Hernquist, L. 2005a, *MNRAS*, 361, 776
- Springel, V., & Hernquist, L. 2003, *MNRAS*, 339, 312
- Springel, V., Pakmor, R., Pillepich, A., Weinberger, R., Nelson, D., Hernquist, L., Vogelsberger, M., Genel, S., Torrey, P., Marinacci, F., & Naiman, J. 2017, ArXiv e-prints
- Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., & Pearce, F. 2005b, *Nature*, 435, 629
- Springob, C. M., Magoulas, C., Colless, M., Mould, J., Erdoğan, P., Jones, D. H., Lucey, J. R., Campbell, L., & Fluke, C. J. 2014, *MNRAS*, 445, 2677
- Stadel, J. G. 2001, PhD thesis, UNIVERSITY OF WASHINGTON
- Steidel, C. C., Erb, D. K., Shapley, A. E., Pettini, M., Reddy, N., Bogosavljević, M., Rudie, G. C., & Rakic, O. 2010, *ApJ*, 717, 289
- Steinhardt, P. J., & Caldwell, R. R. 1998, in *Astronomical Society of the Pacific Conference Series*, Vol. 151, *Cosmic Microwave Background and Large Scale Structure of the Universe*, ed. Y. I. Byun & K. W. Ng, 13
- Stern, J., Hennawi, J. F., Prochaska, J. X., & Werk, J. K. 2016, *ApJ*, 830, 87
- Stinson, G. S., Bailin, J., Couchman, H., Wadsley, J., Shen, S., Nickerson, S., Brook, C., & Quinn, T. 2010, *MNRAS*, 408, 812
- Stinson, G. S., Brook, C., Prochaska, J. X., Hennawi, J., Shen, S., Wadsley, J., Pontzen, A., Couchman, H. M. P., Quinn, T., Macciò, A. V., & Gibson, B. K. 2012, *MNRAS*, 425, 1270
- Strang, G. 1968, *SIAM Journal on Numerical Analysis*, 5, 506
- Strauss, M. A., Davis, M., Yahil, A., & Huchra, J. P. 1990, *ApJ*, 361, 49
- Suresh, J., Bird, S., Vogelsberger, M., Genel, S., Torrey, P., Sijacki, D., Springel, V., & Hernquist, L. 2015, *MNRAS*, 448, 895
- Taruya, A., Bernardeau, F., Nishimichi, T., & Codis, S. 2012, *Phys. Rev. D*, 86, 103528

- Totani, T., Kawai, N., Kosugi, G., Aoki, K., Yamada, T., Iye, M., Ohta, K., & Hattori, T. 2006, in IAU Joint Discussion, Vol. 7, IAU Joint Discussion, 19
- Trainor, R. F., & Steidel, C. C. 2012, *ApJ*, 752, 39
- Tumlinson, J., Thom, C., Werk, J. K., Prochaska, J. X., Tripp, T. M., Katz, N., Davé, R., Oppenheimer, B. D., Meiring, J. D., Ford, A. B., O'Meara, J. M., Peebles, M. S., Sembach, K. R., & Weinberg, D. H. 2013, *ApJ*, 777, 59
- Turner, M. L., Schaye, J., Crain, R. A., Rudie, G., Steidel, C. C., Strom, A., & Theuns, T. 2017, ArXiv e-prints
- Turner, M. L., Schaye, J., Crain, R. A., Theuns, T., & Wendt, M. 2016, *MNRAS*, 462, 2440
- Turner, M. L., Schaye, J., Steidel, C. C., Rudie, G. C., & Strom, A. L. 2014, *MNRAS*, 445, 794
- Valageas, P. 2001, *A&A*, 379, 8
- Valageas, P., Nishimichi, T., & Taruya, A. 2013, *Phys. Rev. D*, 87, 083522
- Vallinotto, A., Das, S., Spergel, D. N., & Viel, M. 2009, *Physical Review Letters*, 103, 091304
- Vallinotto, A., Viel, M., Das, S., & Spergel, D. N. 2011, *ApJ*, 735, 38
- van de Weygaert, R., & van Kampen, E. 1993, *MNRAS*, 263, 481
- Venkatesan, A., Tumlinson, J., & Shull, J. M. 2003, *ApJ*, 584, 621
- Viel, M., Haehnelt, M. G., Bolton, J. S., Kim, T.-S., Puchwein, E., Nasir, F., & Wakker, B. P. 2017, *MNRAS*, 467, L86
- Viel, M., Haehnelt, M. G., & Springel, V. 2006, *MNRAS*, 367, 1655
- Viel, M., Matarrese, S., Mo, H. J., Theuns, T., & Haehnelt, M. G. 2002, *MNRAS*, 336, 685
- Viel, M., Schaye, J., & Booth, C. M. 2013, *MNRAS*, 429, 1734
- Vogele, M. S., Park, C., Geller, M. J., & Huchra, J. P. 1992, *ApJ*, 391, L5
- Vogelsberger, M., Genel, S., Sijacki, D., Torrey, P., Springel, V., & Hernquist, L. 2013, *MNRAS*, 436, 3031
- Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Sijacki, D., Xu, D., Snyder, G., Bird, S., Nelson, D., & Hernquist, L. 2014a, *Nature*, 509, 177
- Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Sijacki, D., Xu, D., Snyder, G., Nelson, D., & Hernquist, L. 2014b, *MNRAS*, 444, 1518
- Wagoner, R. V. 1967, *ApJ*, 149, 465
- Wang, L., Dutton, A. A., Stinson, G. S., Macciò, A. V., Penzo, C., Kang, X., Keller, B. W., & Wadsley, J. 2015, *MNRAS*, 454, 83

- Webster, A. S. 1977, in IAU Symposium, Vol. 74, Radio Astronomy and Cosmology, ed. D. L. Jauncey, 75
- Weinberg, D. H., Hernsquit, L., Katz, N., Croft, R., & Miralda-Escudé, J. 1997, in Structure and Evolution of the Intergalactic Medium from QSO Absorption Line System, ed. P. Petitjean & S. Charlot, 133
- White, M., Myers, A. D., Ross, N. P., Schlegel, D. J., Hennawi, J. F., Shen, Y., McGreer, I., Strauss, M. A., Bolton, A. S., Bovy, J., Fan, X., Miralda-Escude, J., Palanque-Delabrouille, N., Paris, I., Petitjean, P., Schneider, D. P., Viel, M., Weinberg, D. H., Yèche, C., Zehavi, I., Pan, K., Snedden, S., Bizyaev, D., Brewington, H., Brinkmann, J., Malanushenko, V., Malanushenko, E., Oravetz, D., Simmons, A., Sheldon, A., & Weaver, B. A. 2012, MNRAS, 424, 933
- White, M., Pope, A., Carlson, J., Heitmann, K., Habib, S., Fasel, P., Daniel, D., & Lukic, Z. 2010, ApJ, 713, 383
- White, S. D. M., & Frenk, C. S. 1991, ApJ, 379, 52
- White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341
- Widrow, L. 1993, JRASC, 87, 214
- Woods, R. M., Wadsley, J., Couchman, H. M. P., Stinson, G., & Shen, S. 2014, MNRAS, 442, 732
- Worseck, G., Prochaska, J. X., Hennawi, J. F., & McQuinn, M. 2016, ApJ, 825, 144
- Worseck, G., Prochaska, J. X., O'Meara, J. M., Becker, G. D., Ellison, S. L., Lopez, S., Meiksin, A., Ménard, B., Murphy, M. T., & Fumagalli, M. 2014, MNRAS, 445, 1745
- Wyithe, J. S. B., & Loeb, A. 2003, ApJ, 586, 693
- Xu, G. 1995, ApJS, 98, 355
- York, D. G., Adelman, J., Anderson, Jr., J. E., Anderson, S. F., Annis, J., Bahcall, N. A., Bakken, J. A., Barkhouser, R., Bastian, S., Berman, E., Boroski, W. N., Bracker, S., Briegel, C., Briggs, J. W., Brinkmann, J., Brunner, R., Burles, S., Carey, L., Carr, M. A., Castander, F. J., Chen, B., Colestock, P. L., Connolly, A. J., Crocker, J. H., Csabai, I., Czarapata, P. C., Davis, J. E., Doi, M., Dombek, T., Eisenstein, D., Ellman, N., Elms, B. R., Evans, M. L., Fan, X., Federwitz, G. R., Fiscelli, L., Friedman, S., Frieman, J. A., Fukugita, M., Gillespie, B., Gunn, J. E., Gurbani, V. K., de Haas, E., Haldeman, M., Harris, F. H., Hayes, J., Heckman, T. M., Hennessy, G. S., Hindsley, R. B., Holm, S., Holmgren, D. J., Huang, C.-h., Hull, C., Husby, D., Ichikawa, S.-I., Ichikawa, T., Ivezić, Ž., Kent, S., Kim, R. S. J., Kinney, E., Klaene, M., Kleinman, A. N., Kleinman, S., Knapp, G. R., Korienek, J., Kron, R. G., Kunszt, P. Z., Lamb, D. Q., Lee, B., Leger, R. F., Limmongkol, S., Lindenmeyer, C., Long, D. C., Loomis, C., Loveday, J., Lucinio, R., Lupton, R. H., MacKinnon, B., Mannery, E. J., Mantsch, P. M., Margon, B., McGehee, P., McKay, T. A., Meiksin, A., Merelli, A., Monet, D. G., Munn, J. A., Narayanan, V. K., Nash, T., Neilsen, E., Neswold, R., Newberg, H. J., Nichol, R. C., Nicinski, T., Nonino, M., Okada, N., Okamura, S., Ostriker, J. P., Owen, R., Pauls, A. G., Peoples, J., Peterson, R. L., Petravick, D., Pier, J. R., Pope, A., Pordes, R., Prosapio, A., Rechenmacher, R., Quinn, T. R., Richards, G. T., Richmond, M. W., Rivetta, C. H., Rockosi, C. M.,

- Ruthmansdorfer, K., Sandford, D., Schlegel, D. J., Schneider, D. P., Sekiguchi, M., Sergey, G., Shimasaku, K., Siegmund, W. A., Smee, S., Smith, J. A., Snedden, S., Stone, R., Stoughton, C., Strauss, M. A., Stubbs, C., SubbaRao, M., Szalay, A. S., Szapudi, I., Szokoly, G. P., Thakar, A. R., Tremonti, C., Tucker, D. L., Uomoto, A., Vanden Berk, D., Vogeley, M. S., Waddell, P., Wang, S.-i., Watanabe, M., Weinberg, D. H., Yanny, B., Yasuda, N., & SDSS Collaboration. 2000, *AJ*, 120, 1579
- Yoshida, N., Omukai, K., Hernquist, L., & Abel, T. 2006, *ApJ*, 652, 6
- Yu, J. T., & Peebles, P. J. E. 1969, *ApJ*, 158, 103
- Zahn, O., Mesinger, A., McQuinn, M., Trac, H., Cen, R., & Hernquist, L. E. 2011, *MNRAS*, 414, 727
- Zel'dovich, Y. B. 1970, *A&A*, 5, 84
- Zhang, Y., Anninos, P., Norman, M. L., & Meiksin, A. 1997, *ApJ*, 485, 496
- Zwaan, M. A., van der Hulst, J. M., Briggs, F. H., Verheijen, M. A. W., & Ryan-Weber, E. V. 2005, *MNRAS*, 364, 1467
- Zwicky, F. 1933, *Helvetica Physica Acta*, 6, 110



## *Acknowledgements*

I am greatly thankful to my advisor Prof. Dr. Joseph F. Hennawi for giving me the opportunity to undertake the constantly stimulating and challenging work that resulted in this thesis. I am grateful for his guidance and advice throughout my PhD studies that helped me developing my critical thinking and pursuing fascinating research topics. Thanks to him, I could take advantage of several opportunities, in terms of projects, conferences and collaborations that made me grow as a scientist.

I am sincerely grateful to Prof. Dr. Matthias Bartelmann for letting me join his exciting research at the Institute for Theoretical Astrophysics (ITA) in the last few months of my PhD, allowing me to expand my expertise towards analytical techniques in Cosmology. I thank him for his supervision and interest in my progress. I also acknowledge helpful comments on Chapter 4 of this thesis from his side.

I am deeply thankful to Dr. José Oñorbe for his daily mentoring and unlimited patience throughout my PhD. To a large extent, what I learned about simulations is his merit and he has been of great help, both scientifically and humanly. I have been very lucky that we started and ended our experience at the Max Planck Institute for Astronomy (MPIA) together.

It has been a pleasure to collaborate with Dr. Zarija Lukić, to whom I am thankful for his comments on my manuscripts and for giving me the opportunity to give a talk at the Lawrence Berkeley National Laboratory. His suggestions and criticism have strengthened this thesis. I am also thankful to Dr. Annalisa Pillepich and the members of the ENIGMA Group for helpful comments and stimulating discussions on my work.

An enthusiastic thank you to my Master thesis advisor Prof. Dr. Alan F. Heavens. He kept advising me on the project that I worked on as a continuation of my Master thesis, even if I was not his student anymore. His comments and suggestions have been very helpful for accomplishing my first single-author paper and improved the resulting publication.

I warmly thank Prof. Dr. Volker Springel and Prof. Dr. Joseph F. Hennawi for agreeing to referee my PhD thesis. Together with Prof. Dr. Andrea Macciò, I also thank them for being part of my Thesis Committee, monitoring the progress of my work and giving insightful advice about it. I am thankful to Prof. Dr. Luca Amendola and Prof. Dr. Christian Fendt for being part of the examining committee in the final exam of my PhD. Furthermore, I thank the latter for his help as the coordinator of the IMPRS program, for always being understanding with students and for organizing interesting scientific activities and legendary barbecues.

I thank Dr. José Oñorbe, Alexandra García Villota, Dr. Athanasia Tsatsi, Hector Hiß, Manuel Riener, Matthias Samland, Miriam Keppler and Neven Tomičić for providing

me with comments on the draft of the introduction of this thesis. In particular, I am thankful to Matthias Samland for his detailed feedback and for refining the German version of the abstract. I thank all my officemates from MPIA throughout the past four years and from ITA in the last few months for creating a pleasant working environment.

These years would not have been enjoyable without the presence of the friends that I met in Heidelberg. I want to warmly thank Camilla, Salvo, Gabriele and Chiara for the funny time together, both at MPIA and outside, and for bringing a “bit of Italy” at MPIA. Thank you, Fabrizio, for our dinners, chats and journeys together. Thanks, Ema, for the crazy trip around California. Thank you, Anahí, for your friendship and the invitation to Mexico! A big thank you to Rich and Qian Qian for the long hike in the Black Forest, which we should repeat. I am grateful to Miriam, Mayte, Neven, Matthias, Manuel, Aida and Jorge for laughing together in the office and for interesting conversations at lunch, not to mention the funny time in Heidelberg... and Luxembourg.

Unfortunately, life is not made only of happy moments. I am deeply thankful to my friends for helping me through my hardest times, especially Sassa, Fabrizio, Anahí, Chiara, Marco and Martina. Their empathy and encouragement, together with the vicinity of my parents, have been fundamental to overcome the mental health issues that I encountered during my PhD, while little or no institutional support came from MPIA in this respect.

I am very lucky to have got the chance to meet Alexandra. Thank you for the wonderful time spent together, for being of inspiration when I needed a burst of motivation to pursue my scientific and personal interests, and for your kind hospitality.

Whenever I visited my hometown, I was very happy to meet the friends with whom I attended school and university. Despite the distance, we have always retained the willingness to see one another again, and keep us up to date about our lives. Thank you for being there, also when I have been silent for long periods.

My family has been a fundamental support during these years. In particular, I deeply thank and warmly hug my parents, for their efforts to understand my work and my life, for their patience, as well as for their practical and moral support whenever needed. You are the rock on which I can always hold when the sea becomes stormy.

It would be impossible to acknowledge all people that made my stay in Heidelberg a memorable one. If I forgot anyone, it is not out of impoliteness, but because of page constraints. This thesis does not only represent the end of my PhD, but also of 22 years of formal education. I would like to conclude with a quote from my physics teacher at high school, on her last lecture to our class: “Whenever one begins something, such as school or university, it always seems like there is an infinite amount of time ahead. Yet, it is very much like saying ‘infinite’ in Physics: eventually, it ends”.