



**HAL**  
open science

# One- and three-dimensional measurements of the matter distribution from eBOSS and first DESI Lyman- $\alpha$ forest samples

Corentin Ravoux

► **To cite this version:**

Corentin Ravoux. One- and three-dimensional measurements of the matter distribution from eBOSS and first DESI Lyman- $\alpha$  forest samples. *Cosmology and Extra-Galactic Astrophysics* [astro-ph.CO]. Université Paris-Saclay, 2022. English. NNT : 2022UPASP091 . tel-03865230

**HAL Id: tel-03865230**

**<https://theses.hal.science/tel-03865230>**

Submitted on 22 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

One- and three-dimensional  
measurements of the matter distribution  
from eBOSS and first DESI  
Lyman- $\alpha$  forest samples

*Mesures unidimensionnelle et tridimensionnelle de la distribution de  
la matière à partir des forêts Lyman- $\alpha$  d'eBOSS et de DESI*

**Thèse de doctorat de l'université Paris-Saclay**

École doctorale n°576, Particules, hadrons, énergie et noyau : instrumentation,  
imagerie, cosmos et simulation (PHENIICS)  
Spécialité de doctorat : Astroparticules & Cosmologie  
Graduate School : Physique  
Référent : Faculté des sciences d'Orsay

Thèse préparée dans le **Département de Physique des Particules** (Université  
Paris-Saclay, CEA), sous la direction d'**Éric ARMENGAUD**, Directeur de Recherche, et la  
co-direction de **Nathalie PALANQUE-DELABROUILLE**, Directrice de Recherche

**Thèse soutenue à Paris-Saclay, le 22 Septembre 2022, par**

**Corentin RAVOUX**

**Composition du jury**

<b>Sophie HENROT-VERSILLÉ</b> Directrice de recherche, Université Paris-Saclay, Laboratoire de Physique des 2 Infinis Irène Joliot Curie	Présidente
<b>Stéphanie ESCOFFIER</b> Directrice de recherche, Université Aix-Marseille, Centre de Physique des Particules de Marseille	Rapporteuse & Examinatrice
<b>Christophe PICHON</b> Directeur de recherche, Sorbonne Université, Ins- titut d'Astrophysique de Paris	Rapporteur & Examineur
<b>Andreu FONT-RIBERA</b> Chargé de recherche, Universitat Autònoma de Barcelona, Institut de Física d'Altes Energies	Examineur
<b>Éric ARMENGAUD</b> Directeur de Recherche, Université Paris-Saclay, CEA DPhP	Directeur de thèse

**Titre:** Mesures unidimensionnelle et tridimensionnelle de la distribution de la matière à partir des forêts Lyman- $\alpha$  d'eBOSS et de DESI

**Mots clés:** Forêts Lyman- $\alpha$ , Cosmologie, Spectre de puissance, Tomographie, Vides cosmiques, Simulations.

**Résumé:** La distribution de la matière aux échelles cosmologiques peut être prédite dans le cadre du modèle  $\Lambda$ CDM. Elle dépend, entre autres, des propriétés de la matière noire et des neutrinos. Les grands relevés spectroscopiques, tels que eBOSS et DESI, visent à fournir une carte de cette distribution de matière. Cette thèse est une contribution à l'étude des échantillons de forêt Lyman- $\alpha$  provenant de ces relevés. La forêt Lyman- $\alpha$  mesure l'absorption de la lumière de sources distantes par le milieu intergalactique. A ce jour, c'est la meilleure sonde de la distribution de matière à haut redshift et à petite échelle cosmologique (de l'ordre du mégaparsec). Dans cette thèse, j'utilise certains échantillons de forêt Lyman- $\alpha$  de eBOSS et DESI, associés à des quasars situés à un redshift  $z > 2$ , pour étudier leurs propriétés uni- et tri-dimensionnelles. D'une part, je présente la première mesure du spectre de puissance unidimensionnel de la forêt Lyman- $\alpha$  avec les données DESI. Comme DESI est un nouvel instrument, je me concentre sur les systématiques instrumentales, contribuant ainsi à l'amélioration de la mo-

délisation du bruit de l'instrument. Ce spectre de puissance 1D sera utilisé pour contraindre le modèle  $\Lambda$ CDM, la somme des masses des neutrinos, et les propriétés de la matière noire. D'autre part, cette thèse présente une carte tomographique tridimensionnelle, interpolant les forêts Lyman- $\alpha$  du relevé eBOSS. Avec un volume proche d'un gigaparsec cube, cette carte est la plus grande reconstruction tomographique 3D créée à partir des forêts Lyman- $\alpha$ . Elle retrace les fluctuations de la densité de matière, lissées sur une échelle de 13 mégaparsecs, dans la plage de décalage vers le rouge  $2.1 \leq z \leq 3.2$ . Je développe un algorithme pour rechercher des candidats proto-amas et vides cosmiques à partir de cette carte. Je mesure ensuite la corrélation entre les positions des vides détectés et l'échantillon de forêt Lyman- $\alpha$  utilisé pour construire la carte. Cette mesure fournit la première observation des vitesses particulières du gaz intergalactique (RSD) autour des vides à un redshift aussi élevé. En ajustant un modèle linéaire à cette mesure, je mesure le paramètre associé au RSD  $\beta = 0.52$ .

**Title:** One- and three-dimensional measurements of the matter distribution from eBOSS and first DESI Lyman- $\alpha$  forest samples

**Keywords:** Lyman- $\alpha$  forest, Cosmology, Power spectrum, Tomography, Cosmic voids, Simulations

**Abstract:** The distribution of matter on cosmological scales can be predicted in the framework of the  $\Lambda$ CDM model. It depends, among other things, on the properties of dark matter and neutrinos. Large spectroscopic surveys, such as eBOSS and DESI, aim to provide a map of this matter distribution. This thesis is a contribution to the study of Lyman- $\alpha$  forest samples from these surveys. The Lyman- $\alpha$  forest measures the absorption of light from distant sources by the intergalactic medium. As of now, it is the best probe for the distribution of matter at high redshift and small cosmological scale (of the order of a megaparsec). In this thesis, I use some eBOSS and DESI Lyman- $\alpha$  forest samples associated with quasars located at a redshift  $z > 2$ , to study both their one- and three-dimensional properties. On the one hand, I carry out the first measurement of the one-dimensional power spectrum of Lyman- $\alpha$  forest with DESI data. Since DESI is a new instrument, I evaluate the instrumental systematics, contributing to the improvement of

the instrument's noise modeling. This 1D power spectrum will be used to constrain the  $\Lambda$ CDM model, the sum of neutrino masses, and the properties of dark matter. On the other hand, I create a three-dimensional tomographic map, by interpolating the Lyman- $\alpha$  forests from the eBOSS survey. With a volume approaching one gigaparsec cube, this map is the largest 3D tomographic reconstruction created from a Lyman- $\alpha$  forest sample. It traces the fluctuations of matter density, smoothed on a scale of 13 megaparsecs, in the redshift range  $2.1 \leq z \leq 3.2$ . I develop an algorithm to search for candidate proto-clusters and cosmic voids from this map. I then measure the correlation between the positions of detected voids and the Lyman- $\alpha$  forest sample used to construct the map. This measurement provides the first observation of the peculiar velocities of intergalactic gas (RSD) around voids at such a high redshift. By fitting a linear model to this measurement, I measure the associated RSD-related parameter  $\beta = 0.52$ .

# Table of contents

## Remerciements

## Introduction

<b>1</b>	<b>Cosmology</b>	<b>1</b>
<b>1.1</b>	<b>The homogeneous Universe</b>	<b>4</b>
1.1.1	General relativity for cosmology	4
1.1.2	Dynamics of the Universe	6
1.1.3	Distances in cosmology	9
<b>1.2</b>	<b>The <math>\Lambda</math>CDM model</b>	<b>14</b>
1.2.1	History of the Universe	14
1.2.1.1	Primordial Universe	14
1.2.1.2	Cosmic microwave background	18
1.2.1.3	From dark ages to present time	20
1.2.2	Energy content of the Universe	21
1.2.2.1	Relativistic and near-relativistic species	21
1.2.2.2	Non-relativistic matter	25
1.2.2.3	Dark energy	27
<b>1.3</b>	<b>Perturbation theory and large-scale structures</b>	<b>29</b>
1.3.1	Description of cosmic field perturbations	29
1.3.2	Relativistic perturbation theory	32
1.3.2.1	Linearized perturbations	32
1.3.2.2	Linear power spectrum of matter	34
1.3.3	Late-time, small-scale fluctuations	39
1.3.3.1	Eulerian approach	39
1.3.3.2	Lagrangian approach	45
1.3.4	Impact of neutrino masses and warm dark matter	46
1.3.4.1	Neutrino masses	46
1.3.4.2	Warm dark matter	48
<b>1.4</b>	<b>Summary of cosmological observations</b>	<b>52</b>
	<b>Bibliography</b>	<b>54</b>
<b>2</b>	<b>The Lyman-<math>\alpha</math> forest</b>	<b>59</b>
<b>2.1</b>	<b>Properties of the Lyman-<math>\alpha</math> forest</b>	<b>62</b>
2.1.1	Quasars	62
2.1.2	Lyman- $\alpha$ absorption	65

2.1.2.1	The intergalactic medium . . . . .	65
2.1.2.2	Lyman- $\alpha$ forest structure . . . . .	66
2.1.3	Astrophysical systematics of Lyman- $\alpha$ forest . . . . .	71
<b>2.2</b>	<b>Clustering with the Lyman-<math>\alpha</math> forest</b> . . . . .	<b>74</b>
2.2.1	Lyman- $\alpha$ forest correlations . . . . .	74
2.2.2	Modeling of bias and redshift space distortions . . . . .	75
<b>2.3</b>	<b>Some applications of the Lyman-<math>\alpha</math> forest in cosmology</b> . . . . .	<b>79</b>
2.3.1	Observation types . . . . .	79
2.3.2	Lyman- $\alpha$ forest based measurements . . . . .	81
	<b>Bibliography</b> . . . . .	<b>86</b>
<b>3</b>	<b>Simulating the Lyman-<math>\alpha</math> forest</b>	<b>93</b>
<b>3.1</b>	<b>Log-normal mocks for Lyman-<math>\alpha</math> forest</b> . . . . .	<b>96</b>
3.1.1	Saclay mocks . . . . .	96
3.1.2	DESI-Lite mocks . . . . .	99
3.1.3	Modeling systematics in mocks . . . . .	100
<b>3.2</b>	<b>Cosmological simulations</b> . . . . .	<b>102</b>
3.2.1	Overview . . . . .	102
3.2.2	Simulation pipeline . . . . .	106
3.2.3	Simulating the one-dimensional Lyman- $\alpha$ forest power spectrum . . . . .	108
3.2.4	The three-dimensional power spectrum of the Lyman- $\alpha$ forest . . . . .	112
	<b>Bibliography</b> . . . . .	<b>120</b>
<b>4</b>	<b>The SDSS and DESI multi-object spectrograph surveys</b>	<b>127</b>
<b>4.1</b>	<b>Galaxy surveys</b> . . . . .	<b>130</b>
4.1.1	Generalities . . . . .	130
4.1.2	Properties of observed spectra . . . . .	132
<b>4.2</b>	<b>Sloan Digital Sky Survey</b> . . . . .	<b>133</b>
4.2.1	History of SDSS . . . . .	133
4.2.2	Instrument description . . . . .	134
4.2.3	Spectroscopic pipeline . . . . .	136
<b>4.3</b>	<b>Dark Energy Spectroscopic Instrument</b> . . . . .	<b>138</b>
4.3.1	Survey overview . . . . .	138
4.3.2	Instrument description . . . . .	142
4.3.2.1	Telescope . . . . .	142
4.3.2.2	Spectrographs . . . . .	146
4.3.3	Observing with DESI . . . . .	147
4.3.4	DESI spectroscopic pipeline . . . . .	150
4.3.4.1	Spectroscopic extraction . . . . .	150
4.3.4.2	Evaluating the noise . . . . .	153
4.3.4.3	Classifying spectra and measuring redshifts . . . . .	156
	<b>Bibliography</b> . . . . .	<b>157</b>
<b>5</b>	<b>One-dimensional Lyman-<math>\alpha</math> power spectrum with DESI</b>	<b>161</b>
<b>5.1</b>	<b>One-dimensional power spectrum pipeline</b> . . . . .	<b>164</b>
5.1.1	Catalog creation . . . . .	164

5.1.2	Lyman- $\alpha$ contrast extraction . . . . .	167
5.1.3	Power spectrum estimation . . . . .	168
5.1.4	Parameters used for one-dimensional power spectrum calculation . . . . .	170
<b>5.2</b>	<b>Systematics measurement on DESI data</b> . . . . .	<b>172</b>
5.2.1	Noise estimation . . . . .	172
5.2.2	Side-band power spectrum . . . . .	177
5.2.3	Sky emission lines . . . . .	180
<b>5.3</b>	<b>Characterization on mocks</b> . . . . .	<b>183</b>
5.3.1	Pixel masking . . . . .	183
5.3.2	Continuum fitting . . . . .	185
<b>5.4</b>	<b>One-dimensional Lyman-<math>\alpha</math> power spectrum measurement on DESI data</b>	<b>187</b>
5.4.1	Determination of uncertainties . . . . .	188
5.4.2	Measurement . . . . .	191
5.4.3	Summary and prospects . . . . .	192
	<b>Bibliography</b> . . . . .	<b>194</b>
<b>6</b>	<b>Lyman-<math>\alpha</math> tomography</b>	<b>197</b>
<b>6.1</b>	<b>eBOSS data and tomographic method</b> . . . . .	<b>201</b>
6.1.1	Stripe 82 eBOSS data . . . . .	201
6.1.2	Wiener filter . . . . .	204
<b>6.2</b>	<b>Map-making results</b> . . . . .	<b>208</b>
6.2.1	Mocks . . . . .	208
6.2.2	Matter density field of mocks . . . . .	209
6.2.3	eBOSS data . . . . .	211
<b>6.3</b>	<b>Applications</b> . . . . .	<b>215</b>
6.3.1	Void finder . . . . .	215
6.3.2	Catalog stacking . . . . .	217
	6.3.2.1 Void stack . . . . .	217
	6.3.2.2 QSO stacking . . . . .	220
6.3.3	Matter overdensities . . . . .	221
<b>6.4</b>	<b>DESI Lyman-break galaxies secondary target program</b> . . . . .	<b>223</b>
	<b>Bibliography</b> . . . . .	<b>226</b>
<b>7</b>	<b>The correlation between Lyman-<math>\alpha</math> forest and void positions</b>	<b>231</b>
<b>7.1</b>	<b>Lyman-<math>\alpha</math> void cross-correlation pipeline</b> . . . . .	<b>234</b>
7.1.1	Pipeline description . . . . .	234
7.1.2	Pipeline testing . . . . .	238
<b>7.2</b>	<b>Lyman-<math>\alpha</math> void cross-correlation on mocks</b> . . . . .	<b>243</b>
7.2.1	Mock description . . . . .	243
7.2.2	Matter void cross-correlation on mocks . . . . .	244
7.2.3	Impact of systematic effects on the Lyman- $\alpha$ $\times$ Void cross-correlation . . . . .	247
<b>7.3</b>	<b>Modeling the Lyman-<math>\alpha</math><math>\times</math>Void cross-correlation</b> . . . . .	<b>253</b>
7.3.1	Redshift space distortions in Lyman- $\alpha$ $\times$ Void . . . . .	253
7.3.2	Tomographic effect . . . . .	256
7.3.3	Other tests and model summary . . . . .	260
<b>7.4</b>	<b>Lyman-<math>\alpha</math><math>\times</math>Void cross-correlation in SDSS</b> . . . . .	<b>263</b>

7.4.1	Data measurement and comparison with mocks . . . . .	263
7.4.2	Model fitting . . . . .	265
	<b>Bibliography</b> . . . . .	268
	<b>Conclusion and prospects</b>	<b>271</b>
	General conclusion . . . . .	271
	Prospects . . . . .	273
	Bibliography . . . . .	275
	<b>Résumé substantiel</b>	<b>279</b>
<b>I</b>	<b>Le modèle standard de la Cosmologie</b> . . . . .	280
<b>II</b>	<b>La forêt Lyman-<math>\alpha</math></b> . . . . .	283
<b>III</b>	<b>Modélisation numérique de la forêt Lyman-<math>\alpha</math></b> . . . . .	285
III.1	Mocks pour les études des systématiques . . . . .	285
III.2	Simulations cosmologiques . . . . .	287
<b>IV</b>	<b>Spectrographes multi-objets</b> . . . . .	289
<b>V</b>	<b>Le spectre de puissance unidimensionnel de la forêt Lyman-<math>\alpha</math></b> . . . . .	292
<b>VI</b>	<b>Tomographie Lyman-<math>\alpha</math></b> . . . . .	293
<b>VII</b>	<b>Corrélation croisée entre les vides cosmiques et la forêt Lyman-<math>\alpha</math></b> . . . . .	295
	<b>Conclusion</b> . . . . .	297
	<b>Bibliographie</b> . . . . .	298

---

# Remerciements

---

Je tiens tout d'abord à remercier mes directeurs de thèse, Éric Armengaud et Nathalie Palanque-Delabrouille. Vous m'avez tous les deux beaucoup aidé et poussé à donner le meilleur de moi-même. Éric, tu as toujours été présent et tu m'as poussé à améliorer mon esprit d'analyse et ma rigueur scientifique. Tu m'as également aidé à être plus autonome et faire preuve d'initiative. Nathalie, tu as toujours été enthousiaste et positive vis-à-vis de mon travail. Tu as toujours trouvé du temps à m'accorder. Je vous dois beaucoup et j'ai énormément appris à vos côtés. Je remercie également mon "directeur de thèse non officiel", Christophe Yèche, qui m'a énormément aidé dans mon travail et m'a souvent donné des pistes de recherche très fructueuses.

Je remercie les membres de mon jury de thèse pour leur présence lors de ma soutenance. C'est un honneur pour moi de compter dans mon jury de grands experts du domaine. Tout d'abord, je remercie mes deux rapporteurs Stéphanie Escoffier et Christophe Pichon pour avoir pris le temps de lire ma thèse et m'avoir fourni de précieux commentaires qui m'ont permis de l'améliorer. Je remercie également Sophie Henrot-Versillé et Andreu Font-Ribera.

Je remercie tous les membres du groupe de travail Lyman- $\alpha$  de SDSS et DESI avec lesquels j'ai travaillé. Parmi eux, je remercie plus particulièrement Andreu Font-Ribera, Vid Irsic, Julien Guy, Alma González, Naim Karacayli et bien d'autres. Vous m'avez souvent aidé lorsque je me suis retrouvé dans des impasses. Je remercie particulièrement Julian Bautista du CPPM, pour m'avoir guidé dans mon travail et pour m'accueillir en postdoctorat à Marseille. Je remercie également l'ensemble des membres de SDSS et de DESI. Je me suis toujours senti accueilli, écouté et mis en avant, ce qui est plutôt rare dans une collaboration internationale. Je suis heureux de pouvoir continuer à travailler dans cet environnement enrichissant.

Je remercie l'IRFU et plus particulièrement le DPhP ainsi que sa direction Nathalie Besson et Georges Vasseur, pour m'avoir accueilli dans le laboratoire et m'avoir offert un cadre de travail propice au développement personnel et professionnel. Je remercie également Martine Oger qui m'a été d'une aide précieuse pour les démarches administratives. Je remercie également Daniel Pomarède du DEDIP avec qui j'ai eu grand plaisir à collaborer sur la visualisation scientifique de mes données.

Je tiens à remercier très chaleureusement les membres du groupe de Cosmologie du DPhP. Tout d'abord les permanents avec Christophe, Jim, Jean-Marc, Étienne, Vanina, Jean-Batiste et Guillaume. Merci pour votre accueil et votre aide précieuse. J'ai toujours apprécié discuter avec chacun d'entre vous d'un point de vue personnel, mais également d'un point de vue scientifique.

Je remercie également les membres non-permanents du DPhP. Tout d'abord, les postdoctorants Alex et Michael. Michael, tu m'as beaucoup aidé dans mon travail et c'était toujours un plaisir de discuter avec toi. Je remercie mes deux premiers co-bureaux du 41B, Arnaud qui est finalement devenu un permanent durant ma thèse et qui transporte constamment avec lui un optimisme éblouissant partout où il va, et Charles-Antoine avec qui j'ai partagé de très agréables pauses-café tôt le matin et qui maintenant se la coule douce à Hawaii. Je remercie les doctorants qui sont partis du laboratoire, Emmanuel pour sa bonne humeur et ses anecdotes délirantes, Thomas pour toute l'aide qu'il m'a apporté, et Richard pour ses discussions passionnantes sur la vie, la politique et les films. Je tiens à mettre en avant dans cette liste ma collègue Solène qui m'a convaincue de me lancer dans cette thèse, qui m'a toujours aidé et qui m'a toujours tenu au courant des potins du laboratoire. Je remercie mon nouveau co-bureau, Edmond, pour son soutien humoristique et ses gaffes permanentes qui m'impressionneront toujours. Mais aussi Antoine et Alexandre pour leur bonne humeur et leurs blagues pendant les pauses-café. Je remercie également les autres non-permanents du DPhP avec qui j'ai passé des super Friday Lunches. Je souhaite bonne chance aux nouvelles arrivantes, Marie-Lynn et Mathilde pour supporter les trois doctorants restants.

Je tiens à remercier mes amis qui m'ont toujours encouragé en mettant mon travail en valeur, à leur façon. Mes amis de lycée et en particulier Corentin, Matthieu, Arthur et Gabriel, qui m'ont aidé à décompresser et qui m'ont toujours poussé. Mes amis de prépa et plus particulièrement Lucien, Romain, Mathieu, Nicolas, Pauline et Anne-Laurie, qui malgré leurs moqueries, m'ont toujours supporté et soutenu. Je remercie également mes amis de Centrale et en particulier Edmond et Colomban car c'est avec eux que j'ai développé mon goût de la recherche et c'est surtout avec eux que j'ai correctement festoyé. Je remercie également Daniel avec qui j'ai pu partager nos petites galères de doctorants.

Je remercie du fond du cœur l'ensemble de ma famille et de ma belle famille. Tout d'abord ma belle-famille, Claire, Myriam et Serge, pour m'avoir nourris et grandement encouragé. Je remercie mes deux cousins parisiens, Alexis et Valentin, pour les nombreux repas et visionnage de séries les lundis soir. Je remercie énormément mes deux parents Marie-Christine et Claude, et mon frère Damien. Tous les trois vous m'avez poussé à aller plus loin dans mes études et dans mon travail. Vous m'avez toujours encouragé et je vous remercie pour ça. Vous avez toujours dit haut et fort à quel point vous étiez fier de moi, et même si j'ai plus de mal à le dire, je suis tout autant fier de vous.

Finalement, je remercie celle qui m'accompagne et m'aide tous les jours, celle qui partage ma vie depuis maintenant plus de 11 ans, ma compagne Marie-Ange. Tu m'as toujours poussé à donner le maximum de moi-même et tu as toujours été prête à faire de gros sacrifices pour me permettre d'exercer mon métier. Cette thèse, je la dois presque entièrement à toi. Je ne pourrais jamais assez te remercier pour tout ce que tu m'as apporté. Tu n'es pas seulement mon oxygène, tu es tout mon tableau périodique, mon modèle standard et mon Univers tout entier.

---

# Introduction

---

*“Le monde est une belle histoire que chaque génération s’efforce d’améliorer”*

– Georges Lemaître, *L’Hypothèse de l’Atome Primitif*, 1949

IT is not just the world but the whole Universe that cosmologists intend to describe. Achieving this involves observing and simulating this Universe from our little blue marble called Earth. This understanding has been significantly modified over time, going from religious cosmogony to the  $\Lambda$ CDM model (Chap. 1). This Ph.D. manuscript results from my effort to improve the overall understanding of our vast Universe. I completed this work within the "Département de Physique des particules" (DPhP) at the "Commissariat à l’énergie atomique et aux énergies alternatives" (CEA), from 2019 to 2022.

I worked within the SDSS and DESI international collaborations (Chap. 4) in the analysis of the Lyman- $\alpha$  forest (Chap. 2). This observable can be extracted from the measurement of high-redshift quasar spectra and is used to trace the density of neutral hydrogen in the Universe.

Modeling of the Lyman- $\alpha$  forest requires different tools depending on the scale and the type of study considered (Chap. 3). I contributed to creating hydrodynamical simulations for the one- and three-dimensional theoretical modeling of the Lyman- $\alpha$  forest.

The first part of my work consisted in preparing the one-dimensional power spectrum measurement of the Lyman- $\alpha$  forest using the first DESI data, in collaboration with other DESI members (Chap. 5). This measurement is sensitive to instrumental systematics and could only be realized after a thorough examination of their impact.

In a second step, I developed new applications to study the Lyman- $\alpha$  forest in three dimensions. I realized a three-dimensional tomographic map of Lyman- $\alpha$  absorption (Chap. 6) which led to a [first article](#) highlighted in a CEA "[Fait marquant](#)".

Searching for cosmic voids in this tomographic map, I made a correlation measurement between the centers of these voids and the Lyman- $\alpha$  forest (Chap. 7). This measurement led to a [second article](#) which constitutes the first observation of large-scale velocity flows around Lyman- $\alpha$  cosmic voids.

# 1

---

## Cosmology

---

*“L’essentiel est invisible pour les yeux.”*

– Antoine de Saint-Exupéry, *Le Petit Prince*, 1943

---

**Contents**

<b>1.1</b>	<b>The homogeneous Universe</b>	<b>4</b>
1.1.1	General relativity for cosmology	4
1.1.2	Dynamics of the Universe	6
1.1.3	Distances in cosmology	9
<b>1.2</b>	<b>The <math>\Lambda</math>CDM model</b>	<b>14</b>
1.2.1	History of the Universe	14
1.2.1.1	Primordial Universe	14
1.2.1.2	Cosmic microwave background	18
1.2.1.3	From dark ages to present time	20
1.2.2	Energy content of the Universe	21
1.2.2.1	Relativistic and near-relativistic species	21
1.2.2.2	Non-relativistic matter	25
1.2.2.3	Dark energy	27
<b>1.3</b>	<b>Perturbation theory and large-scale structures</b>	<b>29</b>
1.3.1	Description of cosmic field perturbations	29
1.3.2	Relativistic perturbation theory	32
1.3.2.1	Linearized perturbations	32
1.3.2.2	Linear power spectrum of matter	34
1.3.3	Late-time, small-scale fluctuations	39
1.3.3.1	Eulerian approach	39
1.3.3.2	Lagrangian approach	45
1.3.4	Impact of neutrino masses and warm dark matter	46
1.3.4.1	Neutrino masses	46
1.3.4.2	Warm dark matter	48
<b>1.4</b>	<b>Summary of cosmological observations</b>	<b>52</b>
	<b>Bibliography</b>	<b>54</b>

---

---

**C**OSMOLOGY is the science that has the ‘humble’ purpose of describing the evolution and structure of the Universe. This chapter is intended to introduce the concepts necessary to comprehend this science. Throughout history, many different models describing the Universe have been developed. From religious descriptions to the current theories, the vision of the Universe has been greatly modified. This chapter describes the  $\Lambda$ CDM model within the framework of modern cosmology.

The  $\Lambda$ CDM model relies on the theory of general relativity developed by Einstein, Hilbert, Poincaré and Planck. General relativity is the most advanced theory to describe the gravitational force, which plays a major role in the making of the Universe. This theory describes the space-time and it will be the main topic of Sec. 1.1. It provides the tools to model the global evolution of a homogeneous Universe.

The description of the Universe also requires to detail its energetic content. The actual model for the evolution of the Universe is that of the Big Bang. From an extremely hot and dense initial plasma, the Universe expands and cools. Sec. 1.2 is devoted to the description of the Big Bang evolution model, and of the composition of the Universe.

The global evolution of the Universe considers that it is homogeneous and isotropic which is true for the largest scales. However, at the scale of galaxies and large-scale structures of the Universe today, homogeneity and isotropy are not realized at all. The formation of structures comes from density perturbations in the primordial Universe which were amplified by the gravitational force. Sec. 1.3 describes the evolution of these perturbations in the framework of the  $\Lambda$ CDM model and in some extensions, considering massive neutrinos or warm dark matter.

The current cosmological model would not exist without the advances in astronomical observation. Indeed, the discovery of the expansion of the Universe by Hubble and Lemaitre, and then of the acceleration of this expansion thanks to supernovae Ia are striking examples of the need for observation to build the theory. Throughout the chapter, boxes "Feedback from experiments" bring similar examples where experiments have allowed to shine a light on the theory. Sec. 1.4 summarizes the major cosmological probes in use today.

This chapter is based on the reference books [1, 2, 3, 4], on the courses of the master 2 ICFP, on the courses of Euclid Summer School and on the previous theses [5, 6, 7].

## 1.1 The homogeneous Universe

The global description of the Universe does not require to take into account all of the fundamental forces. At large scales, the dominant force is gravity. The best proven theory to date to describe gravity is Einstein's general relativity developed at the beginning of the XX<sup>th</sup> century. This section attempts to introduce the mathematical tools needed for describing the Universe as a whole.

Sec. 1.1.1 describes the main concepts of general relativity useful for cosmology. In particular, it will allow to describe space-time in a general way. Sec. 1.1.2 describes the equations in the particular framework of an isotropic and homogeneous expanding Universe. Finally, in a last Sec. 1.1.3, the mathematical tools necessary to describe cosmological distances and time will be presented.

### 1.1.1 General relativity for cosmology

#### Description of space-time

General relativity is a theoretical framework which allows to link the energy components of the Universe to space-time. To implement it, it is necessary to place oneself in a non-Euclidean manifold which will allow to describe the deformation of space-time.

We place ourselves in the framework of a 4-dimensional Riemannian manifold. This space is defined by giving its **metric tensor**  $g_{\mu\nu}$ . The definition of a metric allows to realize measurements according to the space-time coordinates. In particular, the length element between two events, sometimes called **pseudo-distance**, can be defined by the relation

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu, \quad (1.1)$$

where  $x^\mu$  are the 4-dimensional coordinates of this manifold. We use the convention of Greek letters to represent the coordinates of 4-dimensional vectors (ranging from 0 to 3) and Latin letters for 3-dimensional vectors (ranging from 1 to 3). In this equation and as in the following, the Einstein summation convention is adopted.

For a simple example, to describe the theory of special relativity a 4 dimensional Minkowski space is sufficient. It is defined by the metric

$$g_{\mu\nu} = \eta_{\mu\nu} = \begin{pmatrix} +1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}. \quad (1.2)$$

#### General Relativity

In the case of general relativity, the 4 dimensions of the Riemannian manifold are decomposed into 3 space coordinates and 1 time coordinate.

To reduce the degrees of freedom of a metric, symmetries must be imposed. Einstein formulated the equivalence principle to set the framework of general relativity. This principle can be formulated by:

- **Weak equivalence principle:** The trajectory of a particle or a test body is independent of its composition.

- **Local position invariance:** The result of a non-gravitational experiment is independent of the space-time point considered.
- **Local Lorentz invariance:** The result of a non-gravitational experiment is independent of the free falling referential considered.

It is generally admitted that any theory of gravitation, including general relativity, must verify this principle. In a Lagrangian mechanics context, the **Einstein-Hilbert action**

$$\mathcal{S} = \frac{c^4}{16\pi G_N} \int d^4x \sqrt{-g} (R - 2\Lambda) + \mathcal{S}^{\text{matter}} \quad (1.3)$$

verifies this equivalence principle and is used to describe the evolution of the metric. The coefficient  $\frac{c^4}{16\pi G_N}$  is defined in such a way as to recover Newton's theory of gravity in the classical limit of a Euclidean space-time.  $G_N$  is Newton's constant and  $c$  is the speed of light. Starting from now in this chapter, we use the conventional unit system where  $c = 1$  and the Boltzmann constant  $k_b$  is also equal to unity.

The term  $R$  is the curvature scalar which is related to the Ricci tensor by  $R = g^{\mu\nu} R_{\mu\nu}$ . This tensor is itself a contraction of the Riemann tensor:

$$R_{\mu\nu} = R^{\alpha}_{\mu\alpha\nu}. \quad (1.4)$$

Finally, we can link the Riemann tensor to the metric using Christoffel's symbol  $\Gamma^{\alpha}_{\mu\nu}$ :

$$R^{\mu}_{\nu\alpha\beta} = \partial_{\alpha}\Gamma^{\mu}_{\nu\beta} - \partial_{\beta}\Gamma^{\mu}_{\nu\alpha} + \Gamma^{\mu}_{\sigma\alpha}\Gamma^{\sigma}_{\nu\beta} - \Gamma^{\mu}_{\sigma\beta}\Gamma^{\sigma}_{\nu\alpha}, \quad (1.5)$$

$$\Gamma^{\alpha}_{\mu\nu} = \frac{1}{2}g^{\lambda\alpha} (\partial_{\mu}g_{\lambda\nu} + \partial_{\nu}g_{\mu\lambda} - \partial_{\lambda}g_{\mu\nu}), \quad (1.6)$$

where we define the partial derivative  $\partial_{\mu} = \partial/\partial x^{\mu}$ .

In the first term of the action, the quantity  $\Lambda$  is the **cosmological constant**. It was first introduced by Einstein to reconcile the theory with his vision of a static Universe. After being abandoned by its creator, it was reintroduced in modern cosmology, as it is needed to explain the acceleration of the expansion of the Universe.

The quantities that make up the first term of Eqn. 1.3 are defined from the metric. This part of the action governs the behavior of space-time deformations in an empty region of the Universe. The interaction between space-time and the particle content of the Universe is done by adding the so-called matter action  $\mathcal{S}^{\text{matter}}$ .

Other actions verifying Einstein's equivalence principle can be used to describe a coherent theory of gravity. Maybe the simplest example for such a modified gravity is the  $f(R)$  theory in which the Ricci scalar is replaced by a function of this scalar  $f(R)$ .

The equation governing the evolution of the metric can be obtained by varying the action  $\delta\mathcal{S} = 0$ . This is the **Einstein equation**:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = 8\pi G_N T_{\mu\nu}. \quad (1.7)$$

In this equation, we define the **energy-momentum tensor**, derived from the matter action according to:

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}} \frac{\delta \mathcal{S}^{\text{matter}}}{\delta g^{\mu\nu}}. \quad (1.8)$$

Since this tensor defines the energy content of the Universe, it must be conserved. This is translated in curved space by imposing that its covariant derivative is equal to zero:

$$\nabla_{\alpha} T^{\alpha\mu} = \partial_{\alpha} T^{\alpha\mu} + \Gamma_{\alpha\lambda}^{\alpha} T^{\lambda\mu} + \Gamma_{\alpha\lambda}^{\mu} T^{\alpha\lambda} = 0. \quad (1.9)$$

## 1.1.2 Dynamics of the Universe

### Friedmann-Lemaître-Robertson-Walker metric

General relativity is a framework to describe gravitation. To apply it, it is necessary to define more precisely the system to be studied. In cosmology, the global description of the Universe is realized thanks to a metric that verifies additional symmetries. In particular, it is generally accepted that the Universe is spatially uniform and isotropic at large scales. This statement corresponds to the **cosmological principle**. It is important to note that this strong principle defines the structure of the entire Universe. A more modest principle called Copernican gives the same properties to the observable Universe.

Assuming the cosmological principle, it can be shown that the Universe can be described by the **Friedmann-Lemaître-Robertson-Walker metric** (FLRW)

$$ds^2 = -dt^2 + a^2(t) \left[ \frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2) \right]. \quad (1.10)$$

This metric is defined by three spherical spatial coordinates  $(r, \theta, \phi)$  and a temporal coordinate called cosmic time  $t$ . The sign of the **curvature**  $K$  defines the geometry of the Universe. If  $K$  is strictly negative the Universe is open, if it is strictly positive the Universe is closed and finally flat if  $K = 0$ . The scale factor  $a(t)$  defines the normalized size of the Universe. This parameter controls the contraction or expansion of the Universe. The scale factor at the present time  $(t_0)$  is noted  $a_0 = a(t_0)$ .

It is sometimes useful to consider homogeneous units with respect to the scale factor for both time and space. For this, we define the conformal time by  $d\eta = dt/a(t)$ . By using the conformal time, it is possible to place all the variables in Eqn. 1.10 in the parenthesis. The derivation of a quantity  $X$  with respect to the cosmic time will be noted  $\dot{X}$  and with respect to the conformal time  $X'$ .

### Friedmann equation

To go further and compute the evolution of the scale factor  $a(t)$ , it is necessary to express Einstein's Eqn. 1.7 in accordance with this metric. For this, we start by deriving the curvature scalar according to the FLRW metric:

$$R = 6 \left( \frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{K}{a^2} \right). \quad (1.11)$$

Thus, an extensive calculation provides the terms of the tensor  $G_{\mu\nu}$  appearing in the Einstein equation:

$$\begin{aligned}
G_{00} &= 3 \left( \frac{\dot{a}^2}{a^2} + \frac{K}{a^2} \right), \\
G_{0i} &= 0, \\
G_{ij} &= - \left( \frac{2\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{K}{a^2} \right) g_{ij}.
\end{aligned} \tag{1.12}$$

Considering the imposed space-time symmetries, the most general form of the energy-momentum tensor is the one of a perfect fluid. This tensor can be decomposed into different species  $(a)$ , each characterized by a pressure term  $P^{(a)}$  and a density term  $\rho^{(a)}$ :

$$T_{\mu\nu} = \sum_{(a)} T_{\mu\nu}^{(a)} = \sum_{(a)} (\rho^{(a)} + P^{(a)}) u_\mu u_\nu + P^{(a)} g_{\mu\nu}, \tag{1.13}$$

where  $\mathbf{u}$  is the four-velocity field. Using Einstein's Eqn. 1.7, an extensive calculation yields the two independent **Friedmann equations**:

$$H^2 + \frac{K}{a^2} = \frac{\rho}{3M_{\text{pl}}^2} + \frac{\Lambda}{3}, \tag{1.14}$$

$$\frac{\ddot{a}}{a} = -\frac{1}{6M_{\text{pl}}^2}(\rho + 3P) + \frac{\Lambda}{3}, \tag{1.15}$$

where  $H = \frac{\dot{a}}{a}$  is the **Hubble parameter**,  $M_{\text{pl}} = \frac{1}{\sqrt{8\pi G_{\text{N}}}}$  the reduced Planck mass,  $P = \sum_{(a)} P^{(a)}$  the total pressure and  $\rho = \sum_{(a)} \rho^{(a)}$  the total density. We can also define the Hubble parameter with respect to conformal time  $\mathcal{H} = \frac{a'}{a} = aH$ , where  $a'$  is the derivative of  $a$  with respect to conformal time  $\eta$ .

An additional equation can be derived thanks to the conservation of the energy-momentum tensor given by Eqn. 1.9. Using the FLRW metric, a complete calculation yields the following **conservation equation**:

$$\dot{\rho} + 3H(\rho + P) = 0. \tag{1.16}$$

This equation describes the dilution of particle's energy density by the Hubble flow. It can also be derived from the Friedmann Eqn. 1.14 and 1.15. The system of equations is therefore composed of two independent equations. Its closure requires to specify an equation of state for each species:

$$P^{(a)} = w^{(a)} \rho^{(a)}. \tag{1.17}$$

### Basics of thermodynamics in an expanding Universe

The concepts of pressure and density can be defined by describing the species involved in cosmological scales by a fluid in statistical physics. In particular, a thermodynamical description of fluids in an expanding space explains Eqn. 1.13.

To describe the thermodynamical properties of a cosmological species, we need to define its **distribution function**  $f(\mathbf{x}, \mathbf{p})$ , with  $\mathbf{x}$  the 4-dimensional position and  $\mathbf{p}$  the 4-dimensional

momentum. This function is defined as the density of particles in an infinitesimal volume of space and momentum, normalized by this volume. At thermodynamical equilibrium, it can be expressed by a Fermi-Dirac distribution

$$f_{\text{eq}}(\mathbf{x}, \mathbf{p}) = \frac{1}{\exp[(E(\mathbf{p}) - \mu)/T] \pm 1}, \quad (1.18)$$

where the sign depends on the bosonic or fermionic nature of the particle considered, and  $\mathbf{p} = (p^0, \vec{p})$ . As for the rest of this thesis, we choose the convention of representing 4-dimensional in bold  $\mathbf{X}$  and its 3-dimensional spatial part as  $\vec{X}$ . By integrating the distribution function in momentum space, we can then define thermodynamical quantities such as the number density  $n$ , the energy density  $\rho$ , the pressure  $P$ , and the energy-momentum tensor  $T_{\mu\nu}$ :

$$\begin{aligned} n &= \frac{g}{(2\pi)^3} \int f \sqrt{-g} \, d^3\vec{p}, \\ \rho &= \frac{g}{(2\pi)^3} \int |p^0| f \sqrt{-g} \, d^3\vec{p} = T_{00}, \\ P &= \frac{g}{3(2\pi)^3} \int \frac{|\vec{p}|^2}{|p^0|} f \sqrt{-g} \, d^3\vec{p} = T_{\mu\mu}, \\ T^{\mu\nu} &= \frac{g}{(2\pi)^3} \int p^\mu p^\nu f \frac{\sqrt{-g}}{|p^0|} \, d^3\vec{p}, \end{aligned} \quad (1.19)$$

where  $g$  is the number of degree of freedom.

For a thermally distributed species, applying Eqn. 1.18 to the defined thermodynamical quantities, a detailed calculation allows to simplify these terms in the case of a relativistic species, i.e. massless particles or massive particles with  $T \gg m$  ( $\mu = 0$ ):

$$\begin{aligned} n_{\text{r}} &= \frac{\zeta(3)}{\pi^2} g T^3 \times \begin{cases} 1, & \text{for bosons,} \\ \frac{3}{4}, & \text{for fermions,} \end{cases} \\ \rho_{\text{r}} &= \frac{\pi^2}{30} g T^4 \times \begin{cases} 1, & \text{for bosons,} \\ \frac{7}{8}, & \text{for fermions,} \end{cases} \\ P_{\text{r}} &= \frac{\rho_{\text{r}}}{3}. \end{aligned} \quad (1.20)$$

In the case of a non-relativistic species ( $T \ll m$ )

$$\begin{aligned} n_{\text{nr}} &\simeq g \left( \frac{mT}{2\pi} \right)^{\frac{3}{2}} e^{-\frac{(m-\mu)}{T}}, \\ \rho_{\text{nr}} &= m n_{\text{nr}}, \\ P_{\text{nr}} &= \rho_{\text{nr}} \frac{T}{m} = \rho_{\text{nr}} \frac{\langle v_{\text{th}}^2 \rangle}{3}. \end{aligned} \quad (1.21)$$

Therefore, the equation of state for a relativistic species is  $w = 1/3$  and  $w = \frac{\langle v_{\text{th}}^2 \rangle}{3} \simeq 0$  for non-relativistic matter. From now on, we will refer to non-relativistic species directly as matter and relativistic species as radiation.

### Solutions of the Friedmann equations

Getting back to the Friedmann equations, in the simple case of a single fluid, a short calculation yields the solution of the conservation Eqn. 1.16 for a given equation of state:

Fluid	Equation of state $w$	Energy density $\rho$	Scale factor $a$
Cosmological constant	-1	$\propto \text{constant}$	$\propto e^{Ht}$
Curvature	-1/3	$\propto a^{-2}$	$\propto t$
Non-relativistic matter	0	$\propto a^{-3}$	$\propto t^{2/3}$
Relativistic species	1/3	$\propto a^{-4}$	$\propto t^{1/2}$

TABLE 1.1 – Solutions of Friedmann equations considering different type of fluids.

$$\rho = \rho_p \left( \frac{a_p}{a} \right)^{3(1+w)}, \quad (1.22)$$

where the index  $p$  stands for a particular moment in cosmic time.

The cosmological constant and curvature terms may be included in this formalism, by defining the densities and equations of state of these two fluids by:

$$\begin{aligned} \rho_\Lambda &= \Lambda M_{\text{pl}}^2, & w_\Lambda &= -1, \\ \rho_K &= -3K \frac{M_{\text{pl}}^2}{a^2}, & w_K &= -\frac{1}{3}. \end{aligned} \quad (1.23)$$

For a single fluid, from Eqn. 1.14, we simply obtain the following solution for the scaling factor (if  $w \neq -1$ ):

$$a = a_p \left( \frac{t}{t_p} \right)^{\frac{2}{3(1+w)}}. \quad (1.24)$$

Tab. 1.1 summarizes the results.

Considering several fluids at the same time, we can write in a simplified form the first Friedmann Eqn. 1.14:

$$\Omega_\Lambda + \Omega_K + \sum_{(a)} \Omega_{(a)} = 1, \quad (1.25)$$

by defining the **density parameter**  $\Omega_{(a)} = \frac{\rho_{(a)}}{\rho_c}$  and the **critical density**  $\rho_c = 3M_{\text{pl}}^2 H^2$

Using the index 0 to refer to the present time, we can define the density parameter  $\Omega_{(a),0} = \frac{\rho_{(a),0}}{3M_{\text{pl}}^2 H_0^2}$ . The value of the Hubble parameter at present time  $H_0$  is called **Hubble constant**. With these notations, the evolution of the Hubble parameter is given by:

$$\begin{aligned} H^2(z) &= H_0^2 E(a)^2 \\ &= H_0^2 \left[ \sum_{(a)} \Omega_{(a),0} \left( \frac{a}{a_0} \right)^{-3(1+w_{(a)})} + \Omega_{K,0} \left( \frac{a}{a_0} \right)^{-2} + \Omega_{\Lambda,0} \right]. \end{aligned} \quad (1.26)$$

From now on, we adopt the scale factor convention  $a_0 = 1$ .

### 1.1.3 Distances in cosmology

Based on the FLRW metric, we can define in different ways the time and the distances of objects with respect to an observer in the homogeneous Universe. It is therefore important to formalize precisely these quantities. Several conventions are used to name the distances. In this thesis, distances will be named following [8].

## Redshift

An important quantity in cosmology is the **redshift**. It is defined by the ratio between the observed wavelength  $\lambda_o$  and the emission wavelength  $\lambda_e$  of an object:

$$1 + z = \frac{\lambda_o}{\lambda_e} = \frac{\nu_e}{\nu_o}. \quad (1.27)$$

### Feedback from experiments

Lemaitre and Hubble [9, 10] were the first to highlight the expansion of the Universe. They observed and interpreted the fact that almost all the galaxies close to the Milky Way, formerly called extragalactic nebulae, were moving away and therefore had a positive redshift. They found a linear relationship between the galaxy velocity and its distance  $v = H_0 d$  called **Hubble-Lemaitre law**, as shown in Fig. 1.1.

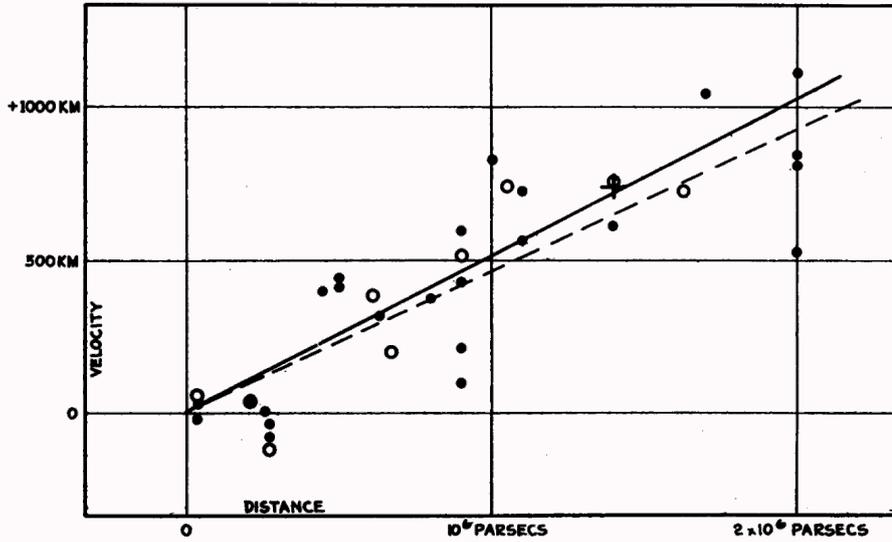


FIGURE 1.1 – Original diagram of the distance and velocity measurements of extragalactic nebulae from [10].

The value of  $H_0$  found by Hubble was  $500 \text{ km}\cdot\text{s}^{-1}\cdot\text{Mpc}^{-1}$ . Today, the measurement of  $H_0$  is an active research topic, and we know that its value is much lower. There is currently a tension between SNIa supernovae ( $73.2 \pm 1.3 \text{ km}\cdot\text{s}^{-1}\cdot\text{Mpc}^{-1}$  in [11]) and the CMB observations ( $67.4 \pm 0.5 \text{ km}\cdot\text{s}^{-1}\cdot\text{Mpc}^{-1}$  in [12]). This tension is the cause of much debate in the scientific community [13, 14].

The redshift is related to the speed at which an object moves away from or towards us. In the Universe, several effects contribute to modify this redshift:

- **Cosmological expansion:** The FLRW metric describes a Universe that can both expand or contract, depending on the sign of the scaling factor derivative. Following Lemaitre and Hubble interpretations, in our Universe  $\dot{a} > 0$ , and the contribution to the redshift of an object due to expansion is driven by the scale factor  $a$ :

$$1 + z_{\text{exp}} = \frac{a_0}{a}.$$

This contribution is determined by considering the path of a photon on a geodesic ( $ds^2 = 0$ ) in the FLRW metric. When only cosmological expansion is considered, the distance of an object can be expressed either by its redshift or by its velocity which is also called **Hubble flow velocity**.

- **Peculiar velocities:** In space-time, the objects also have a peculiar velocity  $v_{\text{pec}}$  on top of the global Hubble flow. It contributes to the redshift by Doppler effect:

$$1 + z_{\text{pec}} = \sqrt{\frac{1 + v_{\text{pec}}/c}{1 - v_{\text{pec}}/c}} \approx 1 + \frac{v_{\text{pec}}}{c}.$$

- **Gravitational redshift:** In the vicinity of massive objects, space-time is distorted. A photon inside a strong gravitational field will lose some energy, and therefore acquire an additional redshift  $z_{\text{grav}}$ , when leaving this gravitational well. In cosmology, this contribution is generally negligible except in some cases (Sachs-Wolfe effect).

By adding these contributions, the redshift  $(1 + z) = (1 + z_{\text{exp}})(1 + z_{\text{pec}})(1 + z_{\text{grav}})$  is used to locate a source in the "redshift space", which is a practical space well adapted to spectroscopic observations.

### Distances

We define the **Hubble distance**  $D_{\text{H}}$  which gives an order of magnitude of the Universe's radius:

$$D_{\text{H}} = \frac{c}{H_0}. \quad (1.28)$$

Due to historical uncertainties related to the measurement of the Hubble constant, we define the **reduced Hubble constant**  $h$  such that  $H_0 = 100h \text{ km}\cdot\text{s}^{-1}\cdot\text{Mpc}^{-1}$ . Distances are often expressed in  $h^{-1}\cdot\text{Mpc}$  (1 parsec = 3.2616 light-years). They correspond to real distances if the Hubble constant is equal to  $100 \text{ km}\cdot\text{s}^{-1}\cdot\text{Mpc}^{-1}$ . The reduced Hubble constant is also used to define **reduced density parameters**  $\omega_{(a),0} = \Omega_{(a),0}h^2$ .

To define the distance of an object from an observer, it is important to take into account the expansion of the Universe. For this reason, the **comoving radial distance** is generally used:

$$D_{\text{C}} = c \int_0^z \frac{dz'}{H(z')}. \quad (1.29)$$

The comoving distance between two objects at the same redshift separated in the sky by an angle  $\delta\theta$  is defined to be  $D_{\text{M}}\delta\theta$  with the **comoving transverse distance**  $D_{\text{M}}$  given by

$$D_{\text{M}} = \begin{cases} D_{\text{H}} \frac{1}{\sqrt{\Omega_k}} \sinh [\sqrt{\Omega_k} D_{\text{C}}/D_{\text{H}}], & \text{for } \Omega_k > 0, \\ D_{\text{C}}, & \text{for } \Omega_k = 0, \\ D_{\text{H}} \frac{1}{\sqrt{|\Omega_k|}} \sin [\sqrt{|\Omega_k|} D_{\text{C}}/D_{\text{H}}], & \text{for } \Omega_k < 0. \end{cases} \quad (1.30)$$

Another distance, practical from an observational point of view, is the **angular diameter distance**  $D_{\text{A}}$ . It is defined as the ratio between the object's physical transverse size and its

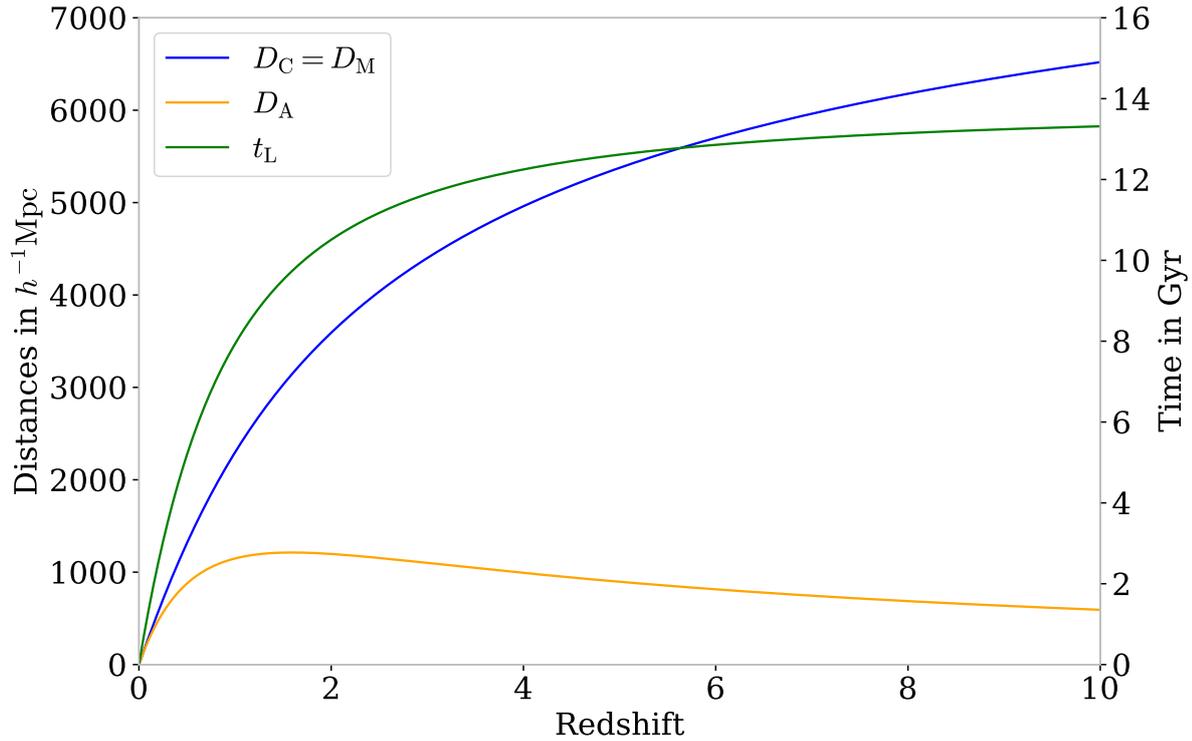


FIGURE 1.2 – Representation of the evolution of the comoving radial distance, the lookback time, the angular diameter distance as a function of redshift. We use a flat  $\Lambda$ CDM model with the Planck 2018 parameters [12]. This graph was created using the python package `cosmoprino`<sup>1</sup> with a CLASS [15, 16] engine.

angular size. It allows to obtain the physical separation between two sources from the separation measured by a telescope. This distance is defined by

$$D_A = \frac{D_M}{1+z}. \quad (1.31)$$

The **luminosity distance**  $D_L$  is used for standard candles. It corresponds to the ratio between the luminosity  $L$  of the source and the received flux  $\phi$ , both integrated over some frequency range:

$$D_L = \sqrt{\frac{L}{4\pi\phi}}. \quad (1.32)$$

It can be related to the comoving transverse distance by the equation

$$D_L = (1+z)D_M. \quad (1.33)$$

Using the previously defined distances, we can define the **comoving volume** at a given redshift by

1. <https://github.com/cosmodesi/cosmoprino>

$$dV_C = D_H \frac{(1+z)^2 D_A^2}{E(z)} d\Omega dz. \quad (1.34)$$

Finally, the **lookback time** corresponds to the difference between the age of the Universe at the time of the emission of a photon by a source located at redshift  $z$  and the age of the Universe at the present time:

$$t_L = \frac{1}{H_0} \int_0^z \frac{dz'}{(1+z') E(z')}. \quad (1.35)$$

All these quantities evolve with the redshift. In Fig. 1.2, the evolution of several distances and of the lookback time as a function of the redshift is shown in the case of a flat Universe with the cosmology given by Planck 2018 [12]. A particularly interesting property of the angular diameter distance is that it reaches a maximum and decreases, meaning that more distant object after  $z \approx 1$  actually appear larger in angular size.

### Angular coordinates

Several spherical coordinate systems can be used to locate a celestial object in astronomy. The definition of these coordinates depends mainly on the chosen reference plane. Once this plane is selected, it defines two hemispheres. One angle defines the projected position along the plane, with respect to an origin, and a second angle defines the rotation outside the plane to the point considered. The horizontal coordinate system defines the plane from the observer's position on Earth. The equatorial coordinate system separates the Earth into two hemispheres from the equatorial plane. Finally, the ecliptic coordinate system is defined with respect to the solar system rotation plane and the galactic coordinate system with respect to the Milky Way rotation plane.

In this thesis, the most used coordinate system is the equatorial one. We will use the **right ascension** (noted  $RA \in [-180, 180]^\circ$ ), which is the angle that gives the rotation on the equatorial plane with an origin taken as its intersection with the ecliptic plane. The **declination** (noted  $DEC \in [-90, 90]^\circ$ ) gives the perpendicular rotation between the observed object and the equator.

## 1.2 The $\Lambda$ CDM model

General relativity and its development in the framework of the FLRW metric offer the necessary tools for the relativistic description of an expanding Universe. We now turn to the description of the Universe's evolution, as well as its content.

As of now, the best description of the Universe is called  $\Lambda$ CDM. This model corresponds to a flat, homogeneous and isotropic Universe, governed by general relativity. In this model, the Universe contains relativistic matter (photons) called radiation, non-relativistic matter (neutrinos, baryons and dark matter) and a cosmological constant. From an initial state, extremely dense and hot, the Universe expands and cools due to this expansion.  $\Lambda$ CDM is the simplest model to explain our current observations. In particular, it allows to make predictions concerning the cosmic microwave background (CMB), the abundance of light elements in the Universe, the presence of large-scale structures that contain galaxies and the recent accelerated expansion of the Universe.

The six parameters of a minimal flat  $\Lambda$ CDM model are

$$\Theta_{\Lambda\text{CDM}} = \{\Omega_{\text{b},0}; \Omega_{\text{cdm},0}; H_0; \tau; A_s; n_s\}.$$

This section will define each of these parameters in relation to the evolution of the Universe. Some additional, "non-minimal" parameters will also be introduced.

Sec. 1.2.1 details the evolution of the Universe in the framework of the Big Bang model. Sec. 1.2.2 focuses on the description of the different constituents which contribute to the energy content of the Universe.

### 1.2.1 History of the Universe

#### 1.2.1.1 Primordial Universe

A simplified timeline marking the different stages of cosmic evolution is shown in Fig. 1.2. Within the **Big Bang** model, the Universe started from an extremely dense and hot state. During the Universe's progressive cooling, the particles of the **standard model of particle physics** combined to form more and more complex systems.

The description of the initial phase following the Big Bang, is very uncertain given the energy scales considered. To describe the physics at the Planck epoch, it would be necessary to have a quantum gravity model sufficiently developed to unify all the fundamental forces.

#### Inflation

At the exit of its initial state, the density matter field of the Universe is not perfectly homogeneous. The current consensus on the origin of these primordial fluctuations is that they would be produced by quantum fluctuations in the Universe in its initial phase.

In addition, there are several issues with the Big Bang model:

- **Horizon problem:** The causality between two events is limited by the speed of light. Consequently, assuming the scale factor evolves as predicted in the case of a radiation domination since  $t = 0$ , the causal horizon on the CMB observed today corresponds to a solid angle of about  $1 \text{ deg}^2$ . Considering this, it is very difficult to explain the homogeneity of the CMB that we observe on the whole sky.

0 s	•	<b>Big Bang</b>
$< 10^{-43}$ s	•	Planck epoch
$< 10^{-36}$ s	•	Grand unification epoch
$< 10^{-32}$ s	•	<b>Inflation</b>
$10^{-32} - 10^{-12}$ s	•	Electroweak epoch
$10^{-12} - 10^{-5}$ s	•	Quark epoch
$10^{-5} - 1$ s	•	Hadron epoch
1 s	•	<b>Neutrino decoupling</b>
1 - 10 s	•	Lepton epoch
10 - $10^3$ s	•	<b>Big Bang nucleosynthesis</b>
10 s - 0.38 Myr	•	Photon epoch
0.38 Myr	•	<b>Recombination (CMB)</b>
0.38 Myr - 150 Myr	•	Dark Ages
150 Myr - 400 Myr	•	Earliest galaxies and stars
250 Myr - 900 Myr	•	<b>Reionization</b>
1 Gyr - 13.8 Gyr	•	<b>Galaxy formation and evolution</b>

TABLE 1.2 – Simplified timeline of the evolution of the Universe.

- **Flatness problem:** This fine-tuning problem is related to the Friedmann equations during a phase of matter and radiation domination. They predict an exponential increase of the curvature. Today most observations tend towards a flat Universe with  $\Omega_{K,0}$  close to zero. To reconcile this observation with the dynamics of the primordial Universe, it is necessary to suppose a strictly zero curvature without any perturbation in the initial state.

It is possible to address these problems by assuming there was a rapid expansion phase in the primordial Universe called **inflation**. This expansion allows both to enlarge the causal horizon to the whole CMB, but also to dampen the exponential increase of the curvature. Solving these two issues respectively constrains the speed and duration of inflation.

Many models were developed to describe inflation. One of the simplest models that addresses most of the problems is the scalar field inflation in the slow-roll regime. However, this scalar field called inflaton is not associated with a field of the standard model.

The inflaton  $\phi$  is defined by its potential  $V(\phi)$ . Minimizing the action associated to inflation gives a Klein-Gordon equation in curved space. Placing ourselves in the flat FLRW metric, we obtain an equation similar to the Friedmann one:

$$H^2 = \frac{1}{3M_{\text{pl}}^2} \left( \frac{1}{2} \dot{\phi}^2 + V(\phi) \right). \quad (1.36)$$

Under certain condition, named "slow-roll", the inflaton field evolves very slowly:

$$\dot{\phi}^2 \ll V(\phi), \quad \ddot{\phi} \ll 3H\dot{\phi}. \quad (1.37)$$

In this regime, Eqn. 1.36 implies that  $H = V(\phi)/(3M_{\text{pl}}^2)$ . Since the inflaton field evolves slowly,  $V(\phi)$  does not vary much in time, and the Hubble parameter is nearly a constant. The situation is similar to the case of a cosmological constant, and the scale factor evolves such that  $a \propto \exp(Ht)$ , hence an inflationary phase.

In addition, thanks to this exponential growth, quantum fluctuations of the inflation field are stretched to macroscopic scales, and are the source of the currently observed inhomogeneities of the Universe.

A complete resolution of inflationary perturbations requires the quantization of the scalar field (see e.g. [1]). An important result of this derivation is that it predicts Gaussian primordial fluctuations, with a dimensionless power spectrum of the spatial curvature perturbations  $\zeta$  given by

$$\frac{k^3}{2\pi^2} P_\zeta(k) = A_s \left( \frac{k}{k_*} \right)^{n_s-1}. \quad (1.38)$$

The subscript  $*$  denotes the choice of a pivot scale  $k_*$  (generally taken equal to 0.05 Mpc). The **spectral amplitude**  $A_s$  and **spectra index**  $n_s$  are related to the inflation parameters in the slow-roll model after a comprehensive derivation:

$$\begin{aligned} n_s &= -M_{\text{pl}}^2 \left( \frac{\partial_\phi V_*}{V_*} \right)^2 - 2M_{\text{pl}} \left[ \frac{\partial_\phi^2 V_*}{V_*} - \left( \frac{\partial_\phi V_*}{V_*} \right)^2 \right] + 1, \\ A_s &= \frac{H_*^2}{4\pi^2 M_{\text{pl}}^4 \left( \frac{\partial_\phi V_*}{V_*} \right)^2}. \end{aligned} \quad (1.39)$$

A term proportional to  $\alpha_s \ln(k/k_*)$  can be added to the spectral index  $n_s$  to parameterize non-standard inflaton models. In the context of single scalar field slow-roll inflation, the running of the spectral index characterized by  $\alpha_s$  is zero. This running is used to put constraints on more advanced inflation models.

This single scalar field model also predicts primordial tensor perturbations, i.e. **primordial gravitational waves** characterized by a tensor-to-scalar ratio  $r$ .

### Universe temperature

After the inflation, the Universe is dominated by radiation. The variation of its scale factor impacts its temperature. Since the radiation density evolution is given by Eqn. 1.20 ( $\rho(a) \propto a^{-4}$ ), we quickly obtain

$$T \propto g_\gamma^{-1/3} a^{-1}, \quad (1.40)$$

where the effective number of degrees of freedom of species in thermal equilibrium with photons is given by

$$g_\gamma(T) = \sum_{(a) = \text{boson}} g_{(a)} \left( \frac{T_{(a)}}{T} \right)^3 + \frac{7}{8} \sum_{(b) = \text{fermion}} g_{(b)} \left( \frac{T_{(b)}}{T} \right)^4. \quad (1.41)$$

In particular,  $g_\gamma(T)$  decreases when a species becomes non relativistic ( $T_{(a)} \leq m_{(a)}$ ).

### Big Bang nucleosynthesis

The complete description of the primordial plasma requires to be placed in the framework of non-equilibrium thermodynamics. An exhaustive description of this theory is beyond the scope of this thesis.

When the Universe expands, interaction processes (reactions) may go out of equilibrium. The moment when a given reaction goes out of equilibrium is called **freeze-out**. To determine the freeze-out time of a reaction, we can compare its reaction rate, noted  $\Gamma$ , to the Hubble constant  $H(t)$ . If  $\Gamma \gg H$ , the reaction is in equilibrium. When the Universe expands,  $H$  decreases but  $\Gamma$  may decrease faster due to the dilution of the interacting species. If  $\Gamma < H$ , the reaction goes out of its equilibrium. In the latter case, the comoving abundance of species intervening in this reaction will be frozen if no other process involve these particles.

At  $T \lesssim 100$  MeV, the primordial plasma is essentially composed of non-relativistic hadrons (protons and neutrons), relativistic leptons (electrons, positrons, and neutrinos) and photons. The thermodynamics of the Universe is maintained at equilibrium by reactions involving weak and electromagnetic interactions:



The ratio between protons and neutrons is then directly given by their equilibrium distribution, i.e. the Boltzmann term:

$$(n/p)_{\text{eq}} = \exp(-(m_n - m_p)/T) \sim 1. \quad (1.43)$$

About 1 second after the Big Bang ( $T \simeq 1$  MeV), the weak interaction rate become too small to maintain equilibrium: there is then a freeze-out of neutrinos. They are decoupled from the primordial plasma and their abundance is fixed.

The primordial plasma equilibrium is modified again between 10 and 1000 s after the Big Bang (temperature between 700 and 50 keV). The decay of the neutron (whose half-life is  $\tau \sim 8$  min), and the production of elements heavier than protons and neutrons by fusion, enter in competition:



In particular, the deuterium production reaction is only possible when the photon density is low enough to prevent the inverse reaction. This phenomenon is called the deuterium bottleneck. It is overcome thanks to the Universe's expansion, which makes the photon density evolve in  $a^{-4}$ . The creation of heavier elements is also facilitated by the annihilation reactions of the remaining positrons with electrons that heat the primordial plasma by the increase of the  $g^\gamma$  number defined by Eqn. 1.41. Other reactions, create a small fraction of elements heavier than  ${}^3\text{He}$ . The main result of these competing processes is that a large amount of helium-4 is created, given by

$$Y_p \simeq \left( \frac{2n}{n+p} \right)_{\text{nuc}} \simeq 0.25. \quad (1.45)$$

The rest of the composition of the Universe is mostly hydrogen (75 %), with traces of deuterium, tritium,  ${}^3\text{He}$ , and  ${}^7\text{Li}$ . The primordial plasma is not hot enough to produce stable

species with more than 7 nucleons. The creation of heavier elements will be done much later, with the stellar nucleosynthesis.

The detailed calculation of these reactions is done with dedicated codes that require input from nuclear physics.

#### Feedback from experiments

The abundances of deuterium in the intergalactic medium (IGM) can be measured by observing absorption features in the spectra of high-redshift quasars [17]. The characteristic absorption line by D could be identified in some specific HCD systems (see Chap. 2), resulting in the measurement  $D/H = 2.527 \times 10^{-5}$  [18]. This is in full agreement with BBN predictions.

### 1.2.1.2 Cosmic microwave background

After BBN, the Universe is still dominated by radiation and appears as a plasma composed of ions (mainly p and He), electrons and photons. The photons are strongly coupled to the electrons. The recombination of atoms is prevented and a permanent photoionization is kept through the following reaction:



Similar reactions also occur for He. During this epoch, the Universe is totally opaque. At about 380,000 years after the Big Bang, the photo-ionization rate becomes lower than  $H$  and the aforementioned equilibrium is broken. The photons decouple from the plasma and the electrons recombine, forming atoms. This **recombination** takes place at redshift  $z_{\text{rec}} = 1090$  and a temperature  $T_{\text{rec}} = 0.3$  eV.

The photons which escape collectively from the primordial plasma constitute the radiation called the **cosmic microwave background (CMB)**. By definition, they are emitted from the last scattering surface.

The CMB is a nearly perfect black-body radiation. The study of its spectral distortions as well as its anisotropies are major sources of information for cosmology. Today, this radiation has undergone a significant redshift and is observed in the microwave electromagnetic domain.

#### Feedback from experiments

CMB was first detected in 1964 by Arno Penzias and Robert Wilson who observed a diffuse background uniformly distributed over the sky. One of the most recent measurements of the CMB was made by the Planck [19] space telescope. Photons from the CMB are observed at different wavelengths in full-sky. From these observations, maps describing the temperature fluctuations of the CMB can be extracted. Similar maps of the polarized components of the CMB can be also obtained.

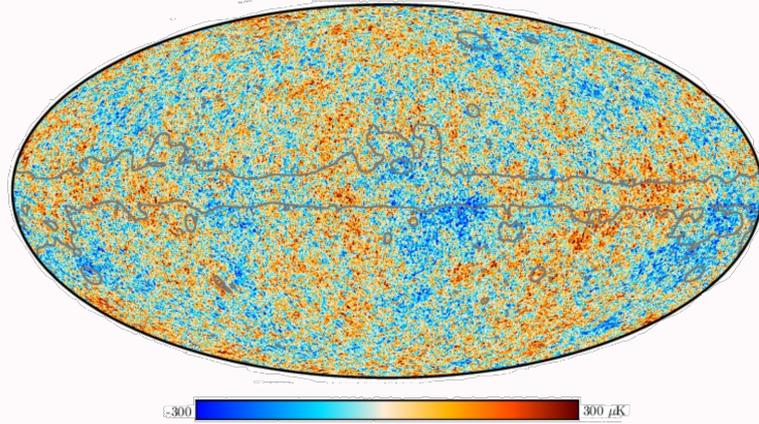


FIGURE 1.3 – CMB temperature map obtained by Planck [19] after foreground subtraction.

The information contained in these anisotropies is compressed through their angular power spectrum. An example of a power spectrum of temperature anisotropies is shown in Fig. 1.4. Angular power spectra can be precisely fit, providing the most significant constraints on key cosmological parameters.

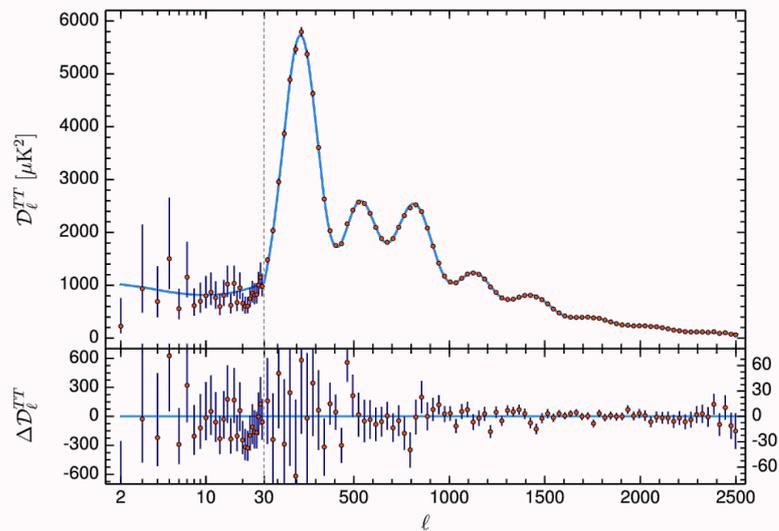


FIGURE 1.4 – Power spectrum of temperature anisotropies obtained by Planck [12]. The value of the x-axis ( $\ell$ ) corresponds to the order in the spherical harmonic decomposition.

An important property of the primordial plasma before recombination is the propagation of acoustic waves, which are called **baryon acoustic oscillations (BAO)**. As baryons and photons are coupled, there is a competition between radiative pressure forces and gravity. This leads to the propagation of acoustic waves.

Shortly after the recombination, at a time called baryons drag  $t_d$ , the pressure of photons is no longer sufficient to prevent the growth of gravitational instabilities of baryons. At this moment, the acoustic waves imprint in the matter distribution a characteristic scale  $r_d \sim 150$  Mpc in comoving space. This scale corresponds to the size of the horizon associated to  $c_s$  at time  $t_d$ .

### Feedback from experiments

The BAO scale  $r_d$  is measured by the relative heights of the CMB peaks in Fig. 1.4. The imprint of BAO on the matter perturbations constitutes a standard ruler. BAO measurement is a key goal of galaxy surveys like SDSS [20, 21, 22, 23, 24] Extended Baryon Oscillation Spectroscopic Survey (eBOSS) [25] or the Dark Energy Spectroscopy survey (DESI) [26, 27] that measure the BAO scale from the positions of galaxies and quasars. The main goal of DESI is to improve the precision of the BAO measurement, covering a wider redshift range. A representation of the BAO peak measurement is given in Fig. 1.5. This peak corresponds to a statistical overdensity between pairs of galaxies separated by a comoving distance  $r_d$ .

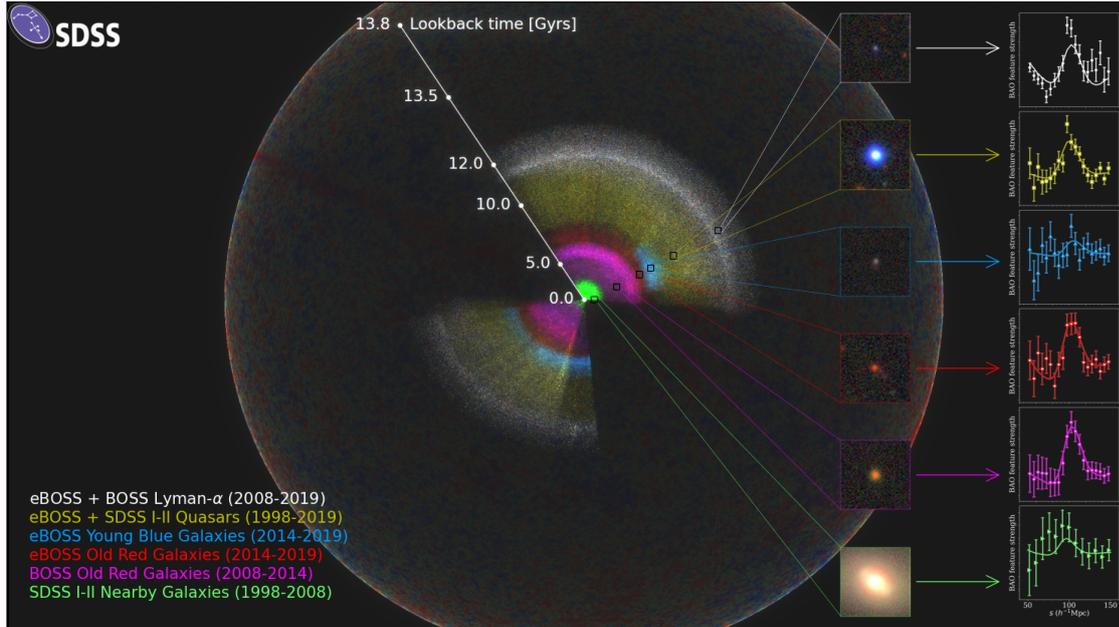


FIGURE 1.5 – Visual representation of the 16<sup>th</sup> data release of the eBOSS survey [25]. For each type of tracers, the BAO peak as seen in the correlation function is represented.

### 1.2.1.3 From dark ages to present time

After the decoupling of photons, the physics describing the Universe drastically changed. From a state of opaque and ionized plasma, the Universe became transparent and neutral.

Following the recombination, during a period of time of about 100 Myr, the Universe was neutral with no other photons than the CMB. The overdensities due to the primordial fluctuations increased by gravitational accretion (see Sec. 1.3). It is estimated that the first galaxies and stars were formed between 150 and 400 Myr after the Big Bang. Due to their primordial composition, these first stars are thought to be very massive, have a very short life span, and emit an important quantity of ionizing UV radiation. This process is believed to be a main cause of the **reionization** of the Universe. After this process, most of the intergalactic medium is ionized, and the Universe is bathed in an ionizing background of UV radiation.

### Feedback from experiments

The reionization history is constrained by the observation of quasar spectra whose Lyman- $\alpha$  forest is totally absorbed by neutral hydrogen [28]. These quasars are located at redshift higher than  $z > 5.5$ . This gives an order of magnitude on the redshift at which the reionization took place.

We can also constrain this reionization process by measuring the optical depth of the CMB from the last scattering surface

$$\tau = \int_0^{z_{\text{rec}}} \frac{n_e \sigma_T}{(1+z)H(z)} dz, \quad (1.47)$$

where  $\sigma_T$  is the Thompson scattering cross section.

The measurement of this optical depth provides the integrated ionization fraction only. Consequently, determining the beginning and duration of reionization remains a challenge of modern cosmology. Furthermore, as the UV background responsible for reionization is produced in compact objects (first stars and quasars), it is inhomogeneous, at least at the beginning.

The last stage of the evolution of the Universe is the evolution of the large structures towards the cosmic web that we observe today. To address mathematically the formation of structures, it is necessary to consider perturbations from a homogeneous and isotropic Universe. We will approach this in more detail in Sec. 1.3.

## 1.2.2 Energy content of the Universe

To complete the description of the  $\Lambda$ CDM model, it is necessary to list the species that compose the Universe and their proportion. In this section, the parameters of the  $\Lambda$ CDM model will be given as inferred from the observation of the CMB by the Planck space telescope [12, 19]. The energy composition of the Universe at present time is represented in Fig. 1.6.

In the present time, most of the energy content of the Universe originates from objects whose nature is still unknown, namely dark energy and dark matter. The evolution of the relative density of each species, as given by the  $\Omega$  parameters, is shown in Fig. 1.7. The early Universe is dominated by radiation. It transitions to a matter-dominated era at the radiation/matter equality characterized by the scale factor  $a_{\text{eq}}$ . After this period during which large structures could form, the Universe entered a phase of accelerated expansion dominated by dark energy.

### 1.2.2.1 Relativistic and near-relativistic species

#### Photons

In the present time, **photons** are a minor species in terms of energy in the  $\Lambda$ CDM model. Stars and galaxies emit photons but they have a negligible weight in this balance. Most of the diffuse background of photons in the Universe come from the CMB (see Sec. 1.2.1.2).

### Feedback from experiments

The best measurement of the CMB temperature is given by COBE/FIRAS  $T_{\text{CMB},0} = 2.7255 \pm 0.0006$  K [30].

Applying Eqn. 1.20, we obtain the following density parameter for photons:

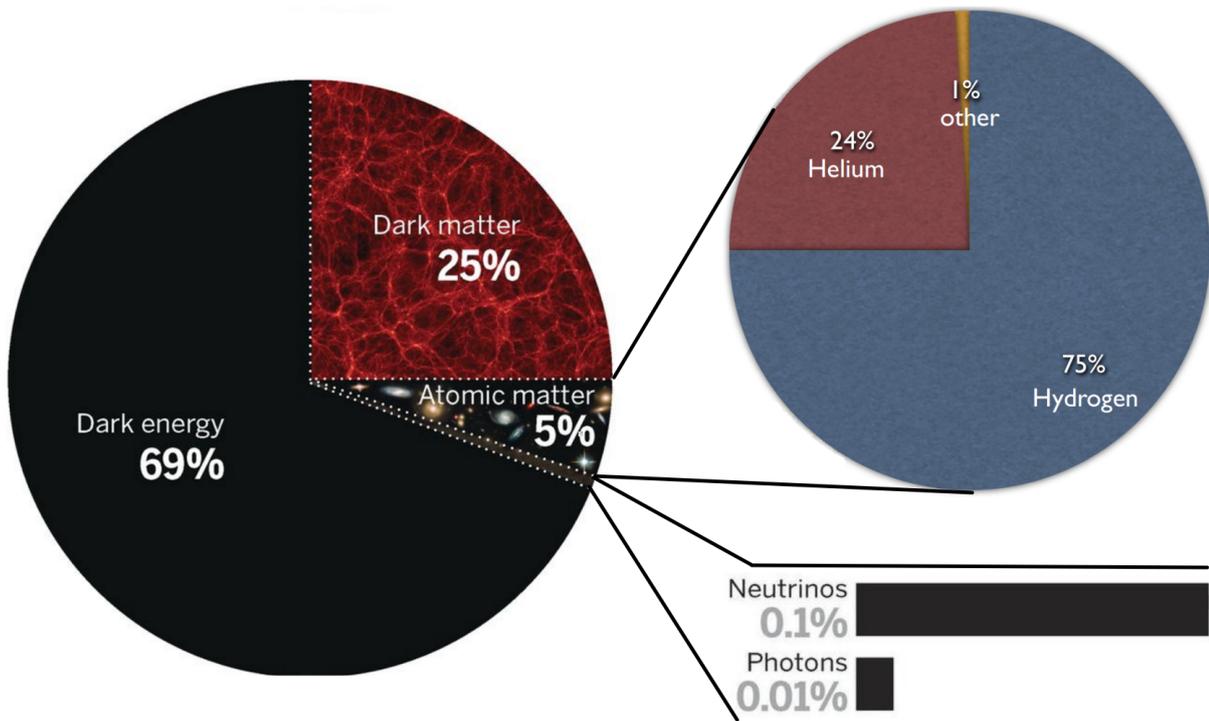


FIGURE 1.6 – Composition of the Universe, in term of energy density as of today. The percentages corresponds approximately to the value given by Planck 2018 cosmology [12]. Figure adapted from [29].

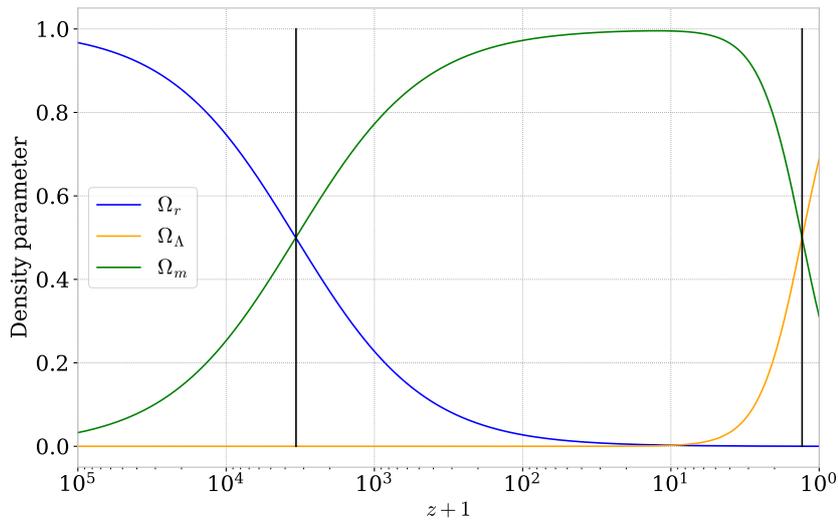


FIGURE 1.7 – Evolution of the density parameters of the radiation, matter and cosmological constant up to today. The black lines separate the three different era. The cosmological model used is a flat  $\Lambda$ CDM model with the Planck 2018 parameters [12]. This graph was created using the python package `cosmoprino`<sup>1</sup>.

$$\Omega_{\gamma,0} = \frac{\rho_{\gamma,0}}{3M_{\text{pl}}^2 H_0^2} = \frac{\pi^2 T_{\text{CMB},0}^4}{30 \cdot 3M_{\text{pl}}^2 H_0^2} = 5.402 \times 10^{-5}. \quad (1.48)$$

## Neutrinos

**Neutrinos** are neutral leptons of the standard model. There are three flavors of neutrinos associated to the three charged leptons. They are called leptonic flavors: electronic ( $\nu_e$ ), muonic ( $\nu_\mu$ ) and tauonic ( $\nu_\tau$ ). The mass of neutrinos is very low compared to the rest of elementary fermion particles in the standard model. The precise determination of their mass is still an unknown in particle physics. In addition, the nature of neutrinos is not yet fully determined. Neutrinos can be their own anti-particles and would then be Majorana particles. Or they could differ from their anti-particle, and be Dirac particles like the charged leptons. Neutrinos interact only through weak interactions and gravity. Given the cross section of the weak interaction, they interact very little with matter. It is the main reason why their observation is very difficult and why there are still many mysteries about neutrinos.

#### Feedback from experiments

The existence of the neutrino was proposed in 1930 by Wolfgang Pauli, to maintain the principle of conservation of energy that the radioactive  $\beta$  decays of atomic nuclei did not seem to respect. Neutrinos were first observed in 1956 by the observation of a neutrino flux from the beta decay reaction in a nuclear reactor [31]. In the first versions of the standard model, neutrinos were considered massless. In the end of the 20<sup>th</sup> century, observations of atmospheric neutrinos confirmed the existence of oscillations between the three neutrino flavors [32]. Solar neutrinos were also used to prove this result. Neutrino oscillations were predicted in 1958 by Bruno Pontecorvo, as a consequence of the fact that neutrinos have a very low but non-zero mass.

The source of neutrinos that interests us in cosmology are those coming from the Big Bang. As we have seen in Sec. 1.2.1.1, the Big Bang model predicts a **cosmic neutrino background (CNB)** composed of neutrino relics decoupled from the primordial plasma shortly before the BBN. The neutrino flux for each species from the CNB as of today is estimated to be about 113 neutrinos and anti-neutrinos per  $\text{cm}^3$ . Since neutrinos decouple from the primordial plasma before the photons, they do not benefit from the heating of the primordial plasma by electron-positron annihilations. A full calculation gives the ratio between the temperatures of neutrinos and photons

$$\frac{T_\nu}{T_\gamma} = \left(\frac{4}{11}\right)^{\frac{1}{3}}. \quad (1.49)$$

The energy density of the CNB depends on whether neutrinos are relativistic or not, i.e. by comparing  $T_\nu$  to  $m_\nu$ . When neutrinos are relativistic, their energy density is then given by

$$\rho_\nu = \rho_\gamma \frac{7}{8} \left(\frac{4}{11}\right)^{4/3} N_{\nu,r}. \quad (1.50)$$

In practice, the decoupling of neutrinos is not instantaneous and a small fraction of the neutrinos recovers energy from the electron-positron annihilation reaction. To take this effect into account, we generally replace the number of massless neutrinos  $N_{\nu,r}$  by an effective number  $N_{\text{eff}}$ . At the time of the CMB, all neutrinos are massless, and we have  $N_{\text{eff}} = 3.046$  [33].  $N_{\text{eff}}$  is best measured by combining CMB and BAO measurements, giving  $N_{\text{eff}} = 2.99 \pm 0.17$  [12].

If neutrinos are relativistic up to today, the neutrino density parameter  $\Omega_{\nu,0}$  is of the same order of magnitude than  $\Omega_{\gamma,0}$  given by Eqn. 1.48.

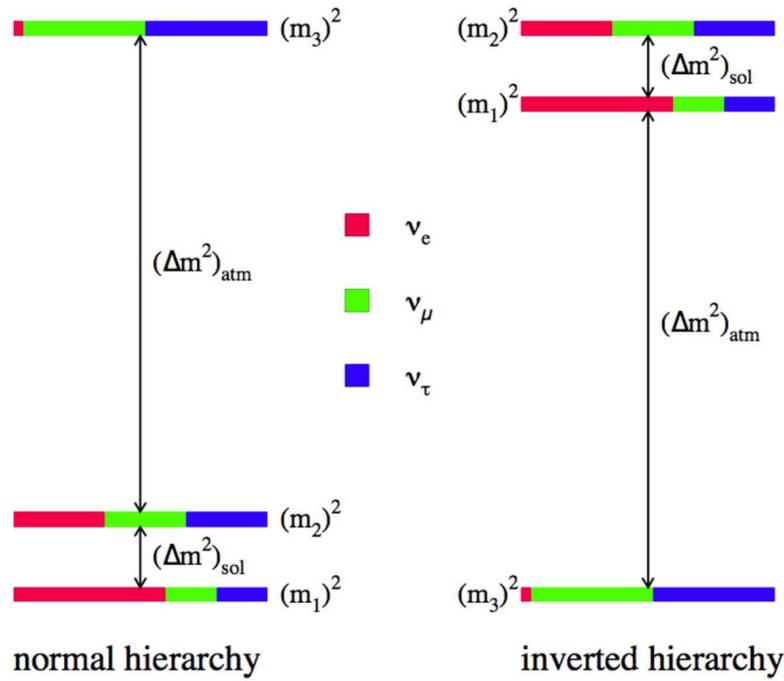


FIGURE 1.8 – Schematic representation of the two different possible hierarchy for individual neutrino masses. Source: wikipedia

### Massive neutrinos

When neutrinos are considered massive, the three neutrino flavor states  $\nu_e$ ,  $\nu_\mu$  and  $\nu_\tau$  are linear combinations of three neutrino mass eigenstates  $\nu_1$ ,  $\nu_2$  and  $\nu_3$ :

$$|\nu_\alpha\rangle = \sum_i U_{\alpha i} |\nu_i\rangle, \quad (1.51)$$

where  $\alpha = (e, \mu, \tau)$ ,  $i = (1, 2, 3)$  and  $U_{\alpha i}$  is the Pontecorvo-Maki-Nakagawa-Sasaka mixing matrix. This matrix has several parameters including mixing angles and a CP violation phase.

Neutrino oscillations from the Sun, nuclear reactors, and particle accelerators allow to measure the mass differences  $\Delta m_{21}^2 = m_2^2 - m_1^2$  and  $|\Delta m_{31}^2| = |m_3^2 - m_1^2|$ . However, there are still unknowns concerning the absolute neutrino mass scale and the mass hierarchy for the different eigenstates. In the context of cosmology, this absolute mass scale is best parameterized by the sum of the neutrino masses  $\sum_\nu m_\nu = m_1 + m_2 + m_3$ . Depending on the sign of  $\Delta m_{31}^2$ , there are two possible ordering. The **normal hierarchy (NH)** with  $\Delta m_{31}^2 > 0$  corresponds to masses  $m_1 < m_2 \ll m_3$  and the **inverted hierarchy (IH)** with  $\Delta m_{31}^2 < 0$  to masses  $m_2 > m_1 \gg m_3$ . A representation of the two possible mass orderings is shown in Fig. 1.8.

Determining the absolute neutrino mass scale nowadays is a major goal for both particle physics and cosmology. This mass scale has an influence in these two fields for the completeness of standard model and the evolution of the Universe.

#### Feedback from experiments

The recent measurements [34] of neutrino oscillations from the Sun and accelerators provides the mass differences at a  $1\sigma$  confidence interval:

$$\begin{cases} \Delta m_{21}^2 = 7.55_{-0.16}^{+0.20} \times 10^{-5} \text{ eV}^2, \\ |\Delta m_{31}^2| = 2.50_{-0.03}^{+0.03} \times 10^{-3} \text{ eV}^2. \end{cases} \quad (1.52)$$

These values provide a constraint on the minimal value of  $\sum_\nu m_\nu$  depending on the mass hierarchy:

$$\begin{cases} \sum_\nu m_\nu > 56 \text{ meV}, & \text{for NH,} \\ \sum_\nu m_\nu > 95 \text{ meV}, & \text{for IH.} \end{cases} \quad (1.53)$$

Considering only results from particle physics, constraints on the absolute mass scale come from neutrinoless double-beta decay (e.g. CUPID [35]) and beta decay of tritium (see e.g. KATRIN [36]) experiments. The best direct measurement comes from KATRIN [37]:

$$m_\beta = \left( \sum_i |U_{ei}|^2 m_i^2 \right)^{1/2} < 0.8 \text{ eV}, \quad \text{for 90\% CL.} \quad (1.54)$$

As a matter of comparison, a cosmological (indirect) constraint on  $\sum_\nu m_\nu$  comes from the combination of the impact of neutrinos on BAO and CMB measurements [38]:

$$\sum_\nu m_\nu < 0.125 \text{ eV}, \quad \text{at 95\% CL.} \quad (1.55)$$

By considering neutrinos as massive, their impact on the energy balance of the Universe is modified. Given our knowledge of the masses, we know that the neutrino density at present time  $\Omega_{\nu,0}$  is completely dominated by the rest energy of non-relativistic neutrinos [33]:

$$\Omega_{\nu,0} = \frac{\sum_\nu m_\nu}{\text{eV}} \frac{1}{93.14 h^2}. \quad (1.56)$$

When the mass of neutrinos is considered non null, we sometimes speak of the  $\Lambda$ CDM $\nu$  model. Massive neutrinos behave like hot dark matter (HDM) because of their weak interaction with baryons and their still large velocity dispersion at current time.

A convenient quantity to determine the proportion of neutrinos in the energy balance is the **neutrino abundance parameter**

$$f_\nu = \frac{\Omega_{\nu,0}}{\Omega_{\text{m},0}}. \quad (1.57)$$

### 1.2.2.2 Non-relativistic matter

The non-relativistic matter during the entire structure formation of the Universe is separated into three parts, the **baryonic matter**, the **cold dark matter (CDM)**, and the non-relativistic (massive) neutrinos mentioned previously:

$$\Omega_{\text{m},0} = \Omega_{\text{b},0} + \Omega_{\text{cdm},0} + \Omega_{\nu,0}. \quad (1.58)$$

#### Baryons

Baryonic matter is composed of ionized or non-ionized atoms and electrons. Including electrons among baryons is a shortcut specific to cosmology since the weight of electrons in the energy balance is low compared to that of nuclei. In the Universe this matter is organized in the form of gas, dust, stars and planets. As seen in Sec. 1.2.1.1, the amount of baryonic matter in

the Universe can be inferred from BBN, but the most precise determination comes from acoustic peaks measured in the CMB anisotropies by Planck [12]:

$$\Omega_{b,0} = 0.0490 \pm 0.0003. \quad (1.59)$$

### Cold Dark Matter

Several astrophysical and cosmological observations indicate today the existence of a matter different from the baryonic matter which plays a primordial role in the formation of the large structures of the Universe. This matter called dark matter (DM) interacts very little with baryonic matter and with itself.

#### Feedback from experiments

One of the first evidence for the existence of dark matter comes from the mismatch between the rotation curve of galaxies and the amount of matter observed. By measuring the velocity profile of a galaxy as shown in Fig. 1.9, it is found that the mass contribution of the visible matter is not sufficient to explain the fast rotation at large radius of the galaxy. Assuming Newtonian dynamics, this observation supports the existence of a dark matter halo around the galaxy.

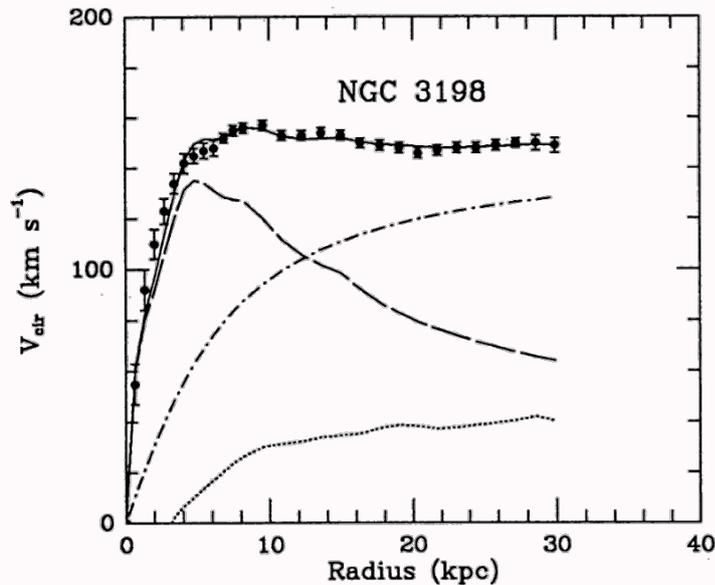


FIGURE 1.9 – Rotation curve of the galaxy NGC 3198 from [39]. The points correspond to the velocity measurement. The dashed line corresponds to the contribution from visible matter and the dotted line from the gas. In order to fit the data points, we need a contribution from a dark matter halo given by the dash-dot curve.

On a larger scale, the observation of a collision of clusters has revealed the existence of a dark matter halo of the size of the galaxy clusters. As shown in the composite image of Fig. 1.10, the baryonic matter of the two clusters, mostly made up of hot X-ray emitting gas in pink, interacted strongly in the form of a shock. In comparison, lensing measurements indicate that most of the mass is located in the blue halo. To understand this mass distribution, it is necessary that there is a matter interacting very little with the baryons and which does not interact through electromagnetic interaction.

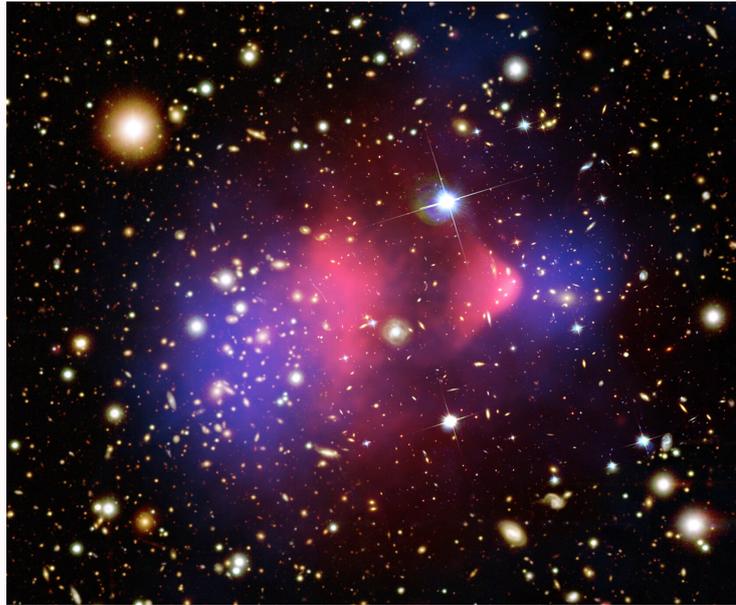


FIGURE 1.10 – Composite image of the bullet cluster. The gas distribution is traced by X-ray (pink). The mass is represented in blue, as obtained from lensing measurement of background galaxies [40].

The existence of dark matter has been demonstrated on many cosmological scales and for very different probes. In particular, the shape of the CMB power spectrum, the explanation of the formation of the current structures, the lensing of halos and many other observations cannot be explained without assuming the existence of dark matter.

The various cosmological probes used to characterize dark matter favor a model of cold (low speed), stable (no decay) and collisionless (weak or zero non-gravitational interactions with itself and baryons) dark matter. This model is called cold dark matter (CDM). The measurement of its density parameter by Planck [12] is the following:

$$\Omega_{\text{cdm},0} = 0.261 \pm 0.002. \quad (1.60)$$

However, dark matter has never been directly observed and its nature is still unknown. Several extensions of the standard model predict the existence of potential candidates such as weakly interacting massive particles (WIMP) motivated by supersymmetry, or a scalar field called axion. Many direct detection experiments of dark matter are looking for the specific signatures of these models with dedicated particle detectors. The association of these searches with cosmological probes already reduces the theoretical space of potential candidates.

### 1.2.2.3 Dark energy

The last component of the  $\Lambda$ CDM model is the one that contributes most in terms of energy at the present time.

#### Feedback from experiments

In the end of the 1990s, measurements of the luminosity distance of type Ia supernovae (SNIa) demonstrated that the Universe is currently undergoing a phase of accelerated

expansion [41, 42]. SNIa are considered as standard candles because their intrinsic luminosity can be normalized. A recent Hubble diagram showing the apparent magnitude of type Ia supernovae is shown in figure 1.11. These observations strongly favor the  $\Lambda$ CDM model. Other major cosmological probes (CMB, BAO) are in agreement with this picture ("concordance cosmological model").

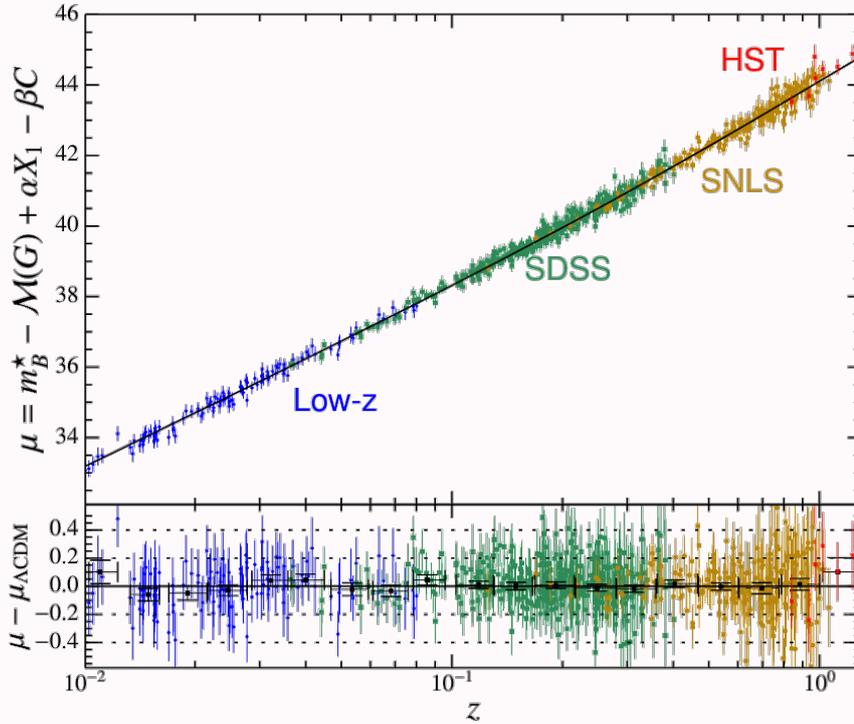


FIGURE 1.11 – Hubble diagram from the SNLS+SDSS sample [43]. The y-axis of the top figure represent the standardized distance modulus  $\mu$  which can be directly linked to the luminosity distance. This diagram is analogous to the original Hubble diagram 1.1.

This accelerated expansion can be explained by a cosmological constant  $\Lambda$  which intervenes in the Friedmann equations as shown in Sec. 1.1. This constant can be considered in the equations as a fluid of density  $\rho_\Lambda = \Lambda M_{\text{pl}}^2$  with an equation of state  $w = -1$ . This fluid is commonly called the **dark energy**. To explore possible deviations from this fixed equation of state, the Chevallier-Polarski-Linder (CPL) parameterization [44, 45] is generally used:

$$w(a) = w_0 + (1 - a)w_a. \quad (1.61)$$

In the case of a cosmological constant model, the constraint given by the CMB observation by Planck [12] is

$$\Omega_{\Lambda,0} = 0.6889 \pm 0.0056. \quad (1.62)$$

The nature of the dark energy is still unknown at present. Many models beyond  $\Lambda$ CDM try to explain this accelerated expansion by introducing a new cosmological fluid (e.g. quintessence, K-essence) or by considering deviations from general relativity over cosmological distances.

## 1.3 Perturbation theory and large-scale structures

Until now, we described the homogeneous (and isotropic) evolution of the Universe. However, this is not the case at all scales. As indicated in Sec. 1.2, there are density perturbations already at the time of recombination. The order of magnitude of these perturbations is given by the CMB temperature fluctuations which are approximately  $\Delta T/T \sim 10^{-5}$ . Later on, these perturbations are amplified, forming the large-scale structures of the Universe.

In this section, we will define the mathematical tools to describe these perturbations in a Universe for which the background evolves according to the Friedmann equation given in Sec. 1.1. This description depends on the scales considered and the amplitude of the perturbations.

First, if we consider scales similar to, or larger than the Hubble radius, it is then necessary to use the framework of general relativity. For smaller scales, a Newtonian description is sufficient. Secondly, if the amplitude of perturbations is small enough ( $(\rho - \bar{\rho})/\bar{\rho} \ll 1$  for density), the equations can be linearized. This amplitude of perturbations depends both on the spatial scale considered and on the epoch. Perturbations are non-linear for small scales and low redshift. For example, to describe the distribution of matter at  $z = 0$ , an order of magnitude of the minimal scale to use linearized equations is  $\sim 50 h^{-1} \cdot \text{Mpc}$ . At the time of the CMB, all observable scales can be considered linear.

All analytical results in this section are based on linearized descriptions. There are analytical theories for going beyond linear theory [2]. However, these methods do not allow to describe strongly non-linear perturbations. In this case, it is necessary to use numerical simulations. Consequently, some equations developed in this section will also be the basis for the description of cosmological simulations which will be presented in Chap. 3.

Sec. 1.3.1 defines the mathematical tools to describe perturbations in general. Then, in Sec. 1.3.2, the equations governing the evolution of perturbations are developed in the framework of general relativity. We will insist on particular aspects of perturbation theory in the Newtonian framework in Sec. 1.3.3. Finally, the impact of neutrinos and warm dark matter models on perturbations will be detailed in Sec. 1.3.4.

### 1.3.1 Description of cosmic field perturbations

It is necessary to define some mathematical tools to describe perturbations. In this section, all the quantities we define depend on the cosmic time  $t$  (or the redshift  $z$ ). For simplification, we will drop this time dependence in the notations, except when necessary.

#### Contrast

If we consider a cosmological quantity (or field)  $X = n, \rho, P, \mathbf{u}$ , we can define its expectation value or ensemble average  $\langle X \rangle$ . In statistical physics, the expectation value corresponds to the average of the quantity  $X$  over several realizations. In cosmology, since there is only one realization of the Universe, we have to assume ergodicity. This means that we approximate the expectation value by the spatial average  $\bar{X} = (1/V) \int_V X(\vec{x}) dV$ .

We can define the **contrast** of a quantity  $X$  at the position  $\vec{x}$  in the following way:

$$\delta_X(\vec{x}) = \frac{X(\vec{x}) - \bar{X}}{\bar{X}}. \quad (1.63)$$

The general expression of the density contrast involves the expectation value, which we have approximated to the spatial average  $\bar{X}$  by ergodicity. By definition, the average value of the contrast is null.

To give an example, the matter density contrast  $\delta_{\rho_m}$  noted  $\delta_m$  is expressed by

$$\delta_m(\vec{x}) = \frac{\rho_m(\vec{x}) - \bar{\rho}_m}{\bar{\rho}_m}. \quad (1.64)$$

The first statistical observable to describe a field  $X$  is the **one-point probability distribution function (PDF)**.

### Two-point correlation function

To describe the characteristic variations of a physical field, the **two-point correlation function** is a key tool:

$$\xi_X(\vec{r}) = \langle \delta_X(\vec{x}) \delta_X(\vec{x} + \vec{r}) \rangle_{\vec{x}}. \quad (1.65)$$

As we have assumed that the Universe is globally homogeneous,  $\xi_X$  does not depend on the position in the Universe  $\vec{x}$ . If we assume the isotropy of the field  $X$ , we obtain that  $\xi_X$  depends only on  $r = |\vec{r}|$ . The two-point correlation function (also called auto-correlation) gives an indication of the correlation of a field with itself for each separation  $r$ . In the particular case of a purely Gaussian field, the 2 point correlation function alone characterizes all the statistical information of a field. All higher order cumulants will be equal to zero.

It is also possible to compute the **two-point cross-correlation function** between two different observables  $X_1$  and  $X_2$ :

$$\xi_{X_1 X_2}(\vec{r}) = \langle \delta_{X_1}(\vec{x}) \delta_{X_2}(\vec{x} + \vec{r}) \rangle_{\vec{x}}. \quad (1.66)$$

### Power spectrum

In cosmology, it is often useful to describe perturbations by modes in Fourier space. The **amplitude of a Fourier mode** is defined by

$$\delta_X(\vec{k}) = \mathcal{F}[\delta_X(\vec{r})] = \int \delta_X(\vec{r}) e^{-i\vec{k}\cdot\vec{r}} d^3\vec{r}. \quad (1.67)$$

For the sake of simplicity, we use in this thesis the  $\vec{k}$ -dependence to indicate that a quantity is expressed in Fourier space. The inverse transform allows us to recover the contrast from the Fourier mode amplitude:

$$\delta_X(\vec{r}) = \mathcal{F}^{-1}[\delta_X(\vec{k})] = \frac{1}{(2\pi)^3} \int \delta_X(\vec{k}) e^{i\vec{k}\cdot\vec{x}} d^3\vec{k}. \quad (1.68)$$

It is then possible to define the **power spectrum** which characterizes the correlation of Fourier modes:

$$\langle \delta_X(\vec{k}) \delta_X(\vec{k}') \rangle = (2\pi)^3 \delta_D^{(3)}(\vec{k} + \vec{k}') P_X(\vec{k}), \quad (1.69)$$

where  $\delta_D^{(3)}$  is the Dirac distribution.

This power spectrum is proportional to the Fourier transform of the two-point correlation function:

$$P_X(\vec{k}) = \frac{1}{(2\pi)^3} \int \xi_X(\vec{r}) e^{-i\vec{k}\cdot\vec{r}} d^3\vec{r}. \quad (1.70)$$

The dimension of the power spectrum is a volume. It is convenient for power spectra comparison to define the dimensionless power spectrum

$$\Delta_X^2(\vec{k}) = \frac{k^3 P_X(\vec{k})}{2\pi^2}. \quad (1.71)$$

When we want to calculate correlations only along one direction, we can use an integral of the power spectrum, called **one-dimensional power spectrum**. This spectrum and its dimensionless definition are then given by:

$$P_{1D,X}(k_{\parallel}) = \int \frac{dk_{\perp}}{(2\pi)^2} P_X(\vec{k}) = \int \frac{dk_{\perp}}{2\pi} k_{\perp} P_X(k_{\parallel}, k_{\perp}), \quad (1.72)$$

$$\Delta_{1D,X}^2(k) = \frac{k P_{1D,X}(k)}{\pi}. \quad (1.73)$$

Similar to the correlation function, it is possible to define the **cross power spectrum**  $P_{X_1 X_2}$  by computing it with two different fields.

### Real space fluctuations

The normalized fluctuations of a quantity  $X$  in a volume  $V$  are characterized by

$$\sigma_X^2(V) = \frac{\langle X(V) - \langle X(V) \rangle \rangle^2}{\langle X(V) \rangle^2}. \quad (1.74)$$

These fluctuations can be related to the integrated power spectrum over a function window  $W$  allowing to limit the integration to the volume  $V$ :

$$\begin{aligned} \sigma_X^2(V) &= \int d^3\vec{k} P_X(\vec{k}) |W_V(\vec{k})|^2, \\ W_V(\vec{k}) &= \frac{1}{V} \int_V d^3\vec{r} e^{i\vec{k}\cdot\vec{r}}. \end{aligned} \quad (1.75)$$

It links the power spectrum to a quantity often used in cosmology, the **amplitude of the matter perturbations** in spheres of  $8 h^{-1} \cdot \text{Mpc}$  comoving radius:

$$\sigma_8(z) = \sqrt{\sigma_m^2(R = 8 h^{-1} \cdot \text{Mpc}, z)}. \quad (1.76)$$

## 1.3.2 Relativistic perturbation theory

To understand large scale perturbations of the Universe, it is necessary to use the framework of general relativity. To define the characteristic scale which defines the need to use relativity, we use the Hubble parameter in conformal time  $\mathcal{H}$ . Comoving Fourier modes with  $k \ll \mathcal{H}$  are called **super-Hubble**, and those with  $k \gg \mathcal{H}$  **sub-Hubble**. The dynamics of the large scale modes (super-Hubble) is governed by general relativity. Conversely, sub-Hubble modes can be described by Newtonian mechanics as shown in Sec. 1.3.3.

The equations of general relativity are non-linear. In order to obtain analytical results on the fluctuations of the large scale metric, it is necessary to linearize the equations. In this section, we will only consider linear perturbations of the metric.

### 1.3.2.1 Linearized perturbations

#### Metric perturbation

The most general **linear perturbation of the FLRW metric** (Eqn. 1.10) is given by

$$ds^2 = a^2(\eta) \left[ -(1 + 2A)d\eta^2 + 2B_i dx^i d\eta + (\gamma_{ij} + h_{ij}) dx^i dx^j \right], \quad (1.77)$$

where  $\gamma_{ij}$  is the spatial part of the FLRW metric:

$$\gamma_{ij} dx^i dx^j = \left[ \frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right]. \quad (1.78)$$

This metric has 10 degrees of freedom (1 for the scalar  $A$ , 3 for the vector  $B_i$  and 6 for the symmetric tensor  $h_{ij}$ ).

It is important to note that a variable  $X$  perturbed in this metric will have Fourier modes  $X(\vec{k})$  which will evolve independently from the others. This property comes from the linearity of perturbations and the homogeneity of the background (and thus from translation invariance).

Using the isotropy of the background, we can use the **scalar-vector-tensor (SVT)** decomposition of the vector  $B_i$  and tensor  $h_{ij}$ :

$$\begin{aligned} B_i &= \partial_i B + \overline{B}_i, \\ h_{ij} &= 2C\gamma_{ij} + \partial_i \partial_j E + \partial_i \overline{E}_j + 2\overline{E}_{ij}. \end{aligned} \quad (1.79)$$

This introduces three new scalars (B,C,E), two vectors  $\overline{B}_i$  and  $\overline{E}_i$  verifying  $\partial_i \overline{B}^i = \partial_i \overline{E}^i = 0$ , and the tensor  $\overline{E}_{ij}$  such that  $\partial_i \overline{E}^{ij} = 0$ .

The metric must also verify gauge invariance, i.e., invariance by change of coordinates  $x^\mu$ . Choosing a gauge and fixing it will remove 4 additional degrees of freedom. Calculations may be realized in a given gauge or by working with gauge-invariant quantities. Here, we describe the gauge-invariant approach.

Two potentials  $\Phi$  and  $\Psi$  called **Bardeen potentials** which are gauge-invariant variables can be defined by

$$\begin{aligned} \Phi &= A + \mathcal{H}(B - E') + (B - E')', \\ \Psi &= -C - \mathcal{H}(B - E'). \end{aligned} \quad (1.80)$$

I recall here that  $'$  corresponds to a derivative with respect to conformal time  $\eta$ . To obtain the dynamics of a perturbed space, we must also parameterize the perturbations of the energy-momentum tensor. They are given by applying the SVT decomposition

$$\begin{aligned}\delta T_{00} &= \rho a^2 (\delta + 2A), \\ \delta T_{0i} &= -\rho a^2 \left[ (1+w)(\partial_i v + \bar{v}_i) + \partial_i B + \bar{B}_i \right], \\ \delta T_{ij} &= P a^2 \left( \gamma_{ij} \frac{\delta P}{P} + h_{ij} + \pi_{ij} \right),\end{aligned}\tag{1.81}$$

where  $P$  is the pressure,  $\rho$  the density and  $w$  the equation of state of unperturbed (homogeneous) quantities of the considered cosmic fluid.

The value  $\delta$  corresponds to density contrast  $\delta_\rho$  and  $\delta P/P$  the pressure fluctuation.  $v$  and  $\bar{v}_i$  are the components of the SVT decomposition of the velocity field. Finally,  $\pi_{ij}$  is the anisotropic energy-momentum tensor and corresponds to the deviation from a perfect fluid behavior.

### Einstein equation

The **perturbed Einstein equation** relates the linear perturbations of the metric to that of the energy-momentum tensor:

$$\delta G_{\mu\nu} = \kappa \delta T_{\mu\nu},\tag{1.82}$$

where  $\kappa = 8\pi G_N = M_{\text{pl}}^{-2}$ . A comprehensive derivation allows us to obtain the perturbed Einstein equations

$$(\Delta + 3K)\Psi = \frac{\kappa}{2} a^2 \rho \delta^C,\tag{1.83}$$

$$\Psi - \Phi = \kappa a^2 P \bar{\pi},\tag{1.84}$$

$$\Psi' + \mathcal{H}\Phi = -\frac{\kappa}{2} a^2 \rho (1+w)V,\tag{1.85}$$

$$\begin{aligned}\Psi'' + (2 + 3c_s^2)\mathcal{H}\Psi' - \left[ (1 + 3c_s^2)K + c_s^2\Delta \right] \Psi \\ = -\mathcal{H}\Phi' - (2\mathcal{H}' + \mathcal{H}^2 + 3c_s^2 K^2)\Phi + P a^2 \kappa \left( \frac{\Delta \bar{\pi}}{3} + \frac{\Gamma}{2} \right).\end{aligned}\tag{1.86}$$

In these equations,  $\delta^C = \delta + \frac{\rho'}{\rho}(v + B)$  is the gauge-invariant density contrast,  $V = v + E'$  is the gauge-invariant velocity scalar and  $K$  is the curvature. The sound speed  $c_s$  and the entropy perturbation  $\Gamma$  are defined by the decomposition of the pressure difference  $\delta P = c_s^2 \delta \rho + P\Gamma$ . Finally the anisotropic stress  $\bar{\pi}$  appears in the SVT decomposition of  $\pi_{ij}$ .

We can also obtain the conservation equation of the fluctuations of the energy-momentum tensor by solving  $\delta T_{\nu;\mu}^\mu = 0$ . After a detailed calculation, we obtain:

$$\begin{aligned}\delta' + 3\mathcal{H} \left( c_s^2 - w \right) \delta = -(1+w) (\Delta V - 3\Psi') - 3\mathcal{H}w\Gamma, \\ V' + \mathcal{H} \left( 1 - 3c_s^2 \right) V = -\Phi - \frac{c_s^2}{1+w} \delta - \frac{w}{1+w} \left[ \Gamma + \frac{2}{3} (\Delta + 3K) \bar{\pi} \right].\end{aligned}\tag{1.87}$$

The Mukhanov-Sasaki variables can be used to obtain the equations in simpler forms:

$$\begin{aligned}
u_s &= \frac{2}{3} \frac{a^2 \Phi}{\mathcal{H}} z, \\
z &= \frac{\mathcal{H}}{a} \left[ \frac{2}{3} (\mathcal{H}^2 + \mathcal{H}' + K) \right]^{(-1/2)}.
\end{aligned} \tag{1.88}$$

Eqn. 1.86 then becomes in Fourier space:

$$u_s'' + \left( k^2 c_s^2 - \frac{z''}{z} \right) u_s = \frac{a^4 \kappa}{3\mathcal{H}} z P \Gamma. \tag{1.89}$$

### 1.3.2.2 Linear power spectrum of matter

#### Composition

The linear perturbed Einstein equation was previously derived for a single cosmic fluid without approximation. To obtain equations adapted to the description of the Universe, we consider two fluids. We simplify this model by considering the case of a flat Universe ( $K = 0$ ) composed only of radiation and dark matter. Furthermore, we consider that these two fluid are perfect, i.e. that  $\Gamma$  and  $\bar{\pi}$  are zero. Using Eqn. 1.84, the two Bardeen potentials  $\Phi$  and  $\Psi$  are equal, and we only consider  $\Phi$  in the following. In addition, the radiation equation of state is constant  $w_r = c_{s,r}^2 = 1/3$ , and as the matter is non relativistic,  $w_m = c_{s,m}^2 = 0$ . The density contrasts of matter and radiation can then be promptly obtained from the conservation Eqn. 1.87:

$$\delta_m' = k^2 V_m + 3\Phi', \tag{1.90}$$

$$\delta_r' = \frac{4}{3} k^2 V_r + 4\Phi', \tag{1.91}$$

$$V_m' = -\mathcal{H} V_m - \Phi, \tag{1.92}$$

$$V_r' = -\Phi - \frac{1}{4} \delta_r. \tag{1.93}$$

This equation system is closed considering the perturbed Einstein equations. By applying Eqn. 1.83 and 1.85 to the density weighted average of the radiation and matter fluid, a detailed calculation yields:

$$-k^2 \Phi = \frac{3}{2} \mathcal{H}^2 \left[ \Omega_m \delta_m + \Omega_r \delta_r - 3\mathcal{H} \left( \Omega_m V_m + \frac{4}{3} \Omega_r V_r \right) \right] = \frac{3}{2} \mathcal{H}^2 \delta^C, \tag{1.94}$$

$$\Phi' + \mathcal{H} \Phi = -\frac{3}{2} \mathcal{H}^2 \left( \Omega_m V_m + \frac{4}{3} \Omega_r V_r \right). \tag{1.95}$$

We use initial conditions such as resulting of the inflation scenario described in Sec. 1.2.1.1, so-called adiabatic fluctuation modes, for which the curvature power spectrum is given by Eqn. 1.38. The corresponding curvature perturbation  $\zeta$  is related to the quantities of the perturbation theory:

$$\zeta = \Phi - \frac{\delta \rho}{[3(\rho + P)]}. \tag{1.96}$$

Adiabatic initial conditions involves that overdensities of all cosmic fluids are in phase. In particular, the relative density contrast  $\delta_m/(1 + w_m) - \delta_r/(1 + w_r)$  is set to zero. Consequently, the relation between matter and radiation contrast is  $\delta_m = (3/4)\delta_r$ . Adiabatic initial conditions

are fixed deep inside the radiation era ( $\Omega_m \ll 1$  and  $\Omega_r \simeq 1$ ) and for super-Hubble modes ( $k \ll \mathcal{H}$ ). Using these approximations on the equation system (1.90 - 1.95), it is possible to link the initial density contrast of matter and radiation to the initial Bardeen potential [1]. Finally, the adiabatic initial conditions read:

$$\Phi'_{\text{ini}} = 0, \quad \delta_{r,\text{ini}} = -2\Phi_{\text{ini}}, \quad \delta_{m,\text{ini}} = -\frac{3}{2}\Phi_{\text{ini}}. \quad (1.97)$$

By considering only radiation ( $w = 1/3$ ) in Eqn. 1.96, the initial Bardeen potential is given by

$$\Phi_{\text{ini}} = \frac{2}{3}\zeta_{\text{ini}}. \quad (1.98)$$

### Super-Hubble modes

We consider super-Hubble modes ( $k \ll \mathcal{H}$ ) during the radiation domination era, and which remain super-Hubble after the transition to the matter era. The evolution of these modes is given by neglecting the  $k^2$  terms in Eqn. 1.90 and 1.91. Using initial conditions defined by Eqn. 1.97, the density contrasts of radiation and matter are quickly expressed by

$$\begin{aligned} \delta_m &= 3\Phi - \frac{9}{2}\Phi_{\text{ini}}, \\ \delta_r &= 4\Phi - 6\Phi_{\text{ini}}. \end{aligned} \quad (1.99)$$

Thus, the evolution of super-Hubble modes is controlled by the Bardeen potential. An exact derivation using Eqn. 1.86 provides the behavior of this potential for both radiation and matter era [1]. Finally, the value of the potential and contrast in the matter domination era are then given by

$$\Phi = \frac{3}{5}\zeta_{\text{ini}}, \quad \delta_r = -\frac{8}{5}\zeta_{\text{ini}}, \quad \delta_m = -\frac{6}{5}\zeta_{\text{ini}}. \quad (1.100)$$

### Sub-Hubble modes

The equation governing the evolution of the matter density contrast can be derived by inserting Eqn. 1.90 and its derivative in Eqn. 1.92:

$$\delta_m'' + \mathcal{H}\delta_m' = 3\Phi'' + 3\mathcal{H}\Phi' - k^2\Phi. \quad (1.101)$$

By combining Eqn. 1.94 and 1.95, the Poisson equation becomes

$$k^2\Phi + 3\mathcal{H}\Phi' + 3\mathcal{H}^2\Phi = -\frac{3}{2}\mathcal{H}^2(\Omega_m\delta_m + \Omega_r\delta_r). \quad (1.102)$$

As shown in [46], before recombination, the fluid fluctuations can be decomposed into slow modes (varying over Hubble time scale  $1/\mathcal{H}$ ) and fast modes (varying over acoustic oscillation time scale  $2\pi/kc_s$ ). The slow mode dominates for dark matter and the fast mode for radiation. Since we are interested in the matter power spectrum here, we only consider slow mode and neglect the  $\delta_r$  term in Eqn. 1.102.

For sub-Hubble modes ( $k \gg \mathcal{H}$ ), one can additionally neglect all the terms which contain the product of  $\mathcal{H}$  with  $\Phi$  and its derivatives with respect to  $k^2\Phi$  in Eqn. 1.101 and 1.102. By combining these two equations, we promptly obtain the Meszaros equation [47] which governs the dark matter perturbations on sub-Hubble scales:

$$\delta_m'' + \mathcal{H}\delta_m' - \frac{3}{2}\mathcal{H}^2\Omega_m\delta_m = 0. \quad (1.103)$$

This equation generally admits two solutions  $D_+$  (growing), and  $D_-$  (decaying) and its complete solution  $\delta_m(\vec{x}, \eta) = D_+(\eta)\delta_+(\vec{x}) + D_-(\eta)\delta_-(\vec{x})$  depends on initial conditions. Both growing and decaying modes depend on cosmological parameters, and we keep these notations in the rest of the thesis to maintain generality.

During the radiation dominated era, using the Friedmann equation, we have  $a \propto \eta$  which implies  $\mathcal{H} = \frac{1}{a}$  and  $\Omega_m \ll 1$ . The equation then becomes

$$\delta_m'' + \frac{1}{a}\delta_m' = 0. \quad (1.104)$$

The growing solution is  $\delta_m \propto \ln(a) + \text{constant}$ . The growth of the fluctuations is slowed down by the fast Hubble expansion driven by radiation.

During the era dominated by matter ( $\Omega_m \simeq 1$ ),  $a \propto \eta^2$  which implies  $\mathcal{H} = \frac{2}{\eta}$ . Eqn. 1.103 then becomes

$$\delta_m'' + \frac{2}{\eta}\delta_m' - \frac{6}{\eta^2}\delta_m = 0. \quad (1.105)$$

In this case, a short calculation shows that  $D_+(\eta) \propto \eta^2$  and  $D_-(\eta) \propto \eta^{-3}$  are both solution of Eqn. 1.105. The growing mode  $D_+(\eta)$  is the solution that gives the gravitational collapse for a Universe dominated by matter.

### Model improvement

To summarize, when we consider only dark matter and radiation, we obtained for the evolution of the matter density contrast during the matter domination phase:

$$\delta_m \propto \Phi_{\text{ini}} \begin{cases} a, & k \gg \mathcal{H} \text{ (sub-Hubble)}, \\ 1, & k \ll \mathcal{H} \text{ (super-Hubble)}. \end{cases} \quad (1.106)$$

To obtain a more accurate evolution of the density contrasts of each species, we must take into account effects that we neglected until now:

- **Baryons:** Before recombination, baryons are non-relativistic like dark matter but they also interact with the photons within the primordial plasma. There are exchanges of momentum between photons and baryons. The main effect is the creation of oscillations of the baryon density contrast due to the propagation of acoustic waves in the primordial plasma. This effect corresponds to the BAO signal presented in Sec. 1.2.1.2. Taking into account this effect requires to model analytically the decoupling between photons and baryons.
- **Dark energy:** To obtain the linear power spectrum at  $z = 0$ , one must also take into account the behavior of the density contrast during the epoch of domination by the cosmological constant.

Full linear calculations taking all these effects into account can be done precisely, by integrating the full set of Boltzmann equations with solvers like **CAMB** [48] or **CLASS** [15, 16]. The contrast information of a species ( $a$ ) at a given redshift is encoded in the **transfer function** defined by

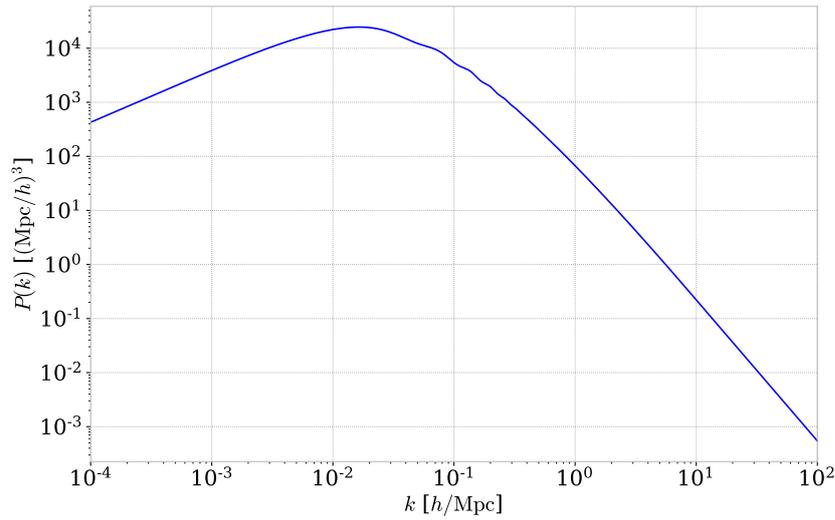


FIGURE 1.12 – Linear matter power spectrum at redshift  $z = 0$ . The cosmological model used is a flat  $\Lambda$ CDM model with the Planck 2018 parameters [12]. This graph was created using the python package `cosmoprime`<sup>1</sup> with a CLASS [15, 16] engine.

$$T_{(a)}(k, z) = \frac{\delta_{(a)}^{\text{C}}(k, z)\delta_{\text{ini}}^{\text{C}}(0)}{\delta_{\text{ini}}^{\text{C}}(k)\delta_{(a)}^{\text{C}}(0, z)}. \quad (1.107)$$

The contrasts of each species are then defined with respect to the initial conditions defined previously.

### Power spectrum

The **linear matter power spectrum** can be expressed from the power spectrum of the curvature perturbations such that

$$P_{\text{m}}(z, k) = (T_{\text{m}}(z, k))^2 P_{\zeta}(k). \quad (1.108)$$

In the  $\Lambda$ CDM model, without massive neutrino, the density contrast of matter can be expressed by

$$\delta_{\text{m}}(z, k) = \frac{\Omega_{\text{b}}\delta_{\text{b}}(z, k) + \Omega_{\text{cdm}}\delta_{\text{cdm}}(z, k)}{\Omega_{\text{b}} + \Omega_{\text{cdm}}}. \quad (1.109)$$

By adding the effects described above and using the initial conditions defined by scalar field inflation detailed in Sec. 1.2.1.1, a comprehensive derivation yields an approximate analytical expression for the late-time, linear power spectrum of matter:

$$P_{\text{m}}(\eta, k) \approx \left(\frac{6}{5}\right)^2 \eta^4 g^2(\eta) A_{\text{s}} k^{n_{\text{s}}} \begin{cases} 1, & k < \mathcal{H}_{\text{eq}}, \\ \left(\frac{k_{\text{eq}}}{k}\right)^4 \left(1 + \ln \frac{k}{k_{\text{eq}}}\right)^2 \times (\text{baryon suppression}), & k > \mathcal{H}_{\text{eq}}, \end{cases} \quad (1.110)$$

where  $\eta$  is the conformal time and the function  $g$  takes into account the evolution of perturbations during  $\Lambda$  domination era.

A numerical calculation of this linear power spectrum at  $z = 0$  performed with the CLASS [15, 16] software is reported in Fig. 1.12.

### Feedback from experiments

If we restrict ourselves to scales for which perturbations are weak, the linear model of Eqn. 1.110 and its numerical calculation in figure 1.12 are valid to describe matter perturbations. The value of the perturbations will typically depend on the scale considered and the redshift. Additionally, one can also find the linear spectrum of matter from a tracer which is in a non-linear regime, by using some approximations. Today, different surveys allow to probe the matter distribution in the Universe. A representation of the linear matter power spectrum inferred from different surveys is given in Fig. 1.13. CMB observations allow to obtain information on the largest scales. Galaxy surveys such as eBOSS [25] or DESI [26, 27] allow to measure the region of BAO peak described in Sec. 1.2.1.2. But these surveys are also sensitive to the distribution of matter on smaller scales, for example with the use of the Lyman- $\alpha$  forest which is the focus of this thesis. Finally, observations of weak lensing of galaxies by matter also provide information on its distribution (e.g. DES [49]).

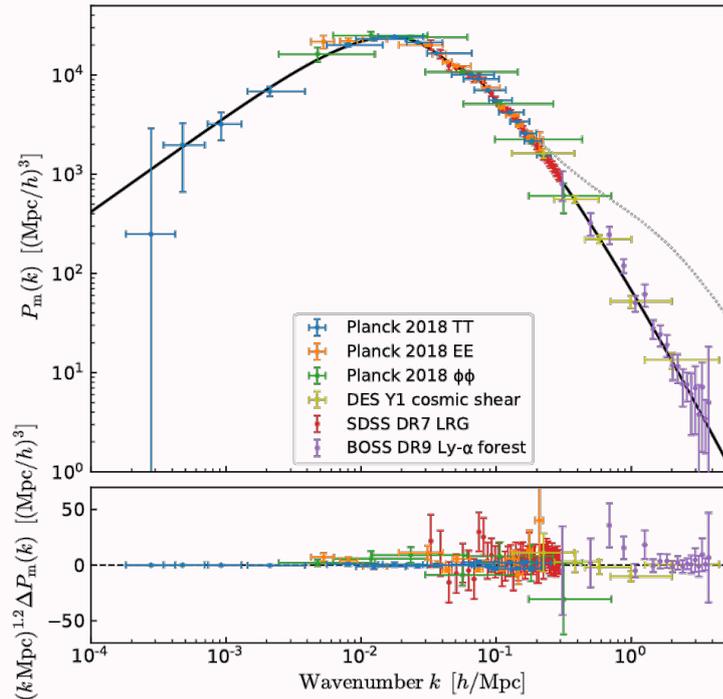


FIGURE 1.13 – Inference of the linear power spectrum of matter at  $z = 0$  from different cosmological probes, taken from [50]. The solid line shows the linear matter power spectrum obtained from CLASS with a  $\Lambda$ CDM model using Planck [12] data. The dotted line shows a non-linear correction to the power spectrum using Halofit.

### 1.3.3 Late-time, small-scale fluctuations

In this section, we will focus in more detail on the theory of matter fluctuations in the sub-Hubble regime, where a development in the Newtonian framework is sufficient.

At some point, the perturbations are too large to be described by linear equations. This happens for small scales and for recent cosmic times. In particular, the non-linear equations developed here will be the basis for the simulations of large structure formation detailed in Chap. 3. To follow the evolution of cosmological fluids in the sub-Hubble regime, two different approaches are possible:

- **Eulerian:** The properties of the fluid are defined at each point in space. We then define fields such as velocity or temperature at each point in space. This approach is similar to the description developed in the framework of general relativity in Sec. 1.3.2.
- **Lagrangian:** The fluid is considered as a system of moving fluid "elements". The Lagrangian approach consists in following the position, the velocity and the different properties of these "particles". The fields are then not described as a function of the position in space but as a function of the initial conditions of the particles.

The equations in the Newtonian framework of each approach will be detailed in Sec. 1.3.3.1 and 1.3.3.2.

#### 1.3.3.1 Eulerian approach

##### Vlasov-Poisson system

As mentioned in Sec. 1.1.2, the description of a cosmic fluid can be done by considering its distribution function  $f(\vec{x}, \vec{p}, t)$ . Using Liouville's theorem and the mean field assumption in phase space, the distribution function is conserved. By applying this theorem we obtain the **Vlasov equation**, translating the dynamics of this function in phase space:

$$\frac{df}{d\eta} = \frac{\partial f}{\partial \eta} + \frac{d\vec{x}}{d\eta} \cdot \frac{\partial f}{\partial \vec{x}} + \frac{d\vec{p}}{d\eta} \cdot \frac{\partial f}{\partial \vec{p}} = 0. \quad (1.111)$$

We place ourselves in the Newtonian framework, within an expanding space characterized as in Sec. 1.1 by the scale factor  $a$ . In this space, it is possible to define **comoving coordinates**. The fields of comoving position  $\vec{x}$ , comoving velocity  $\vec{v}$  and comoving momentum  $\vec{p}$  are defined by

$$\begin{aligned} \vec{x} &= \frac{\vec{r}(t)}{a(t)}, \\ \vec{v} &= \dot{\vec{r}}, \\ \vec{p} &= am\vec{u}, \end{aligned} \quad (1.112)$$

where  $\vec{u} = \dot{\vec{x}}' = a(t)\dot{\vec{x}}$  is the physical velocity.

The comoving quantities are very often used as they allow to describe space-time while "factorizing" the Hubble expansion. The physical speed  $\vec{u}$  can be related to the comoving speed  $\vec{v}$  by

$$\vec{v}(t) = H\vec{r} + \vec{u}. \quad (1.113)$$

We now consider a matter fluid with no other interaction than gravity. Newton's equation governs the dynamics of this system in an expanding space:

$$\frac{d\vec{p}}{d\eta} = -am\nabla_x\Phi(\vec{x}). \quad (1.114)$$

The gravitational potential is determined by the **Poisson equation** in an expanding space

$$\Delta\Phi = \frac{1}{2}a^2M_{\text{pl}}^{-2}\bar{\rho}\delta = \frac{3}{2}\mathcal{H}^2\Omega_m\delta, \quad (1.115)$$

where  $\delta = \delta_\rho$  is the total matter density contrast. The Vlasov equation then becomes:

$$\frac{\partial f}{\partial\eta} + \frac{\vec{p}}{am} \cdot \frac{\partial f}{\partial\vec{x}} - am\nabla_x\Phi(\vec{x}) \cdot \frac{\partial f}{\partial\vec{p}} = 0. \quad (1.116)$$

### Fluid momentum equations

The Vlasov equation involves  $f$  in a 6 dimensions space. To obtain equations involving the moments of the distribution function ( $\rho$ ,  $\rho\vec{u}$ ,  $\rho\vec{u}^2\dots$ ), we integrate the Vlasov equation after having multiplied it by those moments.

The direct integration of the Vlasov equation with respect to  $\vec{p}$  provides after a complete calculation the **continuity equation**

$$\delta' + \frac{1}{\bar{\rho}}\nabla_x \cdot [\rho\vec{u}] = 0, \quad (1.117)$$

where  $'$  corresponds to derivative with respect to conformal time. Similarly, after multiplication by  $\vec{p}$ , the integration gives the **Euler equation**

$$\vec{u}' + \mathcal{H}\vec{u} + \vec{u}\nabla_x \cdot \vec{u} = -\nabla\Phi - \frac{1}{\rho}\nabla_x(\rho\vec{\sigma}), \quad (1.118)$$

with  $\vec{\sigma}$  the stress tensor that characterizes the fluid velocity dispersion:

$$\frac{\int p_i p_j f(\vec{x}, \vec{p}) d^3\vec{p}}{m^2 a^2 \int f(\vec{x}, \vec{p}) d^3\vec{p}} = u_i u_j + \sigma_{ij}. \quad (1.119)$$

The integration according to higher-order moments provides a hierarchy of equations connecting the moments together. The system of equations formed by 1.117, 1.118 and 1.115 is not closed. In order to solve it, it is necessary to make an assumption about the velocity distribution at large orders. In the cases of baryons and cold dark matter, the velocity dispersion on cosmological scales is small, so one can approximate  $\vec{\sigma}$ . For cold dark matter at early times, we have  $\vec{\sigma} = 0$ . This approximation breaks in a non-linear regime when dark matter trajectories intersect for the first time (shell crossing). For baryons, we make the approximation that the fluid is locally thermalized which simplifies the expression of the stress tensor to  $\sigma_{ij} = \delta_{ij} \frac{P}{\rho}$  [2].

In addition, we suppose that the pressure perturbation is related to the density perturbation by the sound velocity  $\delta P = c_s^2 \delta \rho$ . We then quickly obtain the following closed system of equations:

$$\begin{aligned} \delta' + \nabla_x \cdot [(1 + \delta)\vec{u}] &= 0, \\ \vec{u}' + \mathcal{H}\vec{u} + \vec{u}\nabla_x \cdot \vec{u} + c_s^2 \nabla_x \delta + \nabla_x \Phi &= 0, \\ \Delta \Phi - \frac{1}{2} M_{\text{pl}}^{-2} \bar{\rho} a^2 \delta &= 0. \end{aligned} \quad (1.120)$$

### Linearization

Linearization allows us to obtain an analytical solution of this system of equations. At linear order, taking the first equation of 1.120 and the divergence of the second one, we simply obtain:

$$\delta' + \nabla_x \cdot \vec{u} = 0, \quad (1.121)$$

$$(\nabla_x \cdot \vec{u})' + \mathcal{H} \nabla_x \cdot \vec{u} + c_s^2 \Delta \delta + \Delta \Phi = 0. \quad (1.122)$$

Combining this system of equations with the Poisson equation, a complete calculation yields a single second order differential equation:

$$\delta'' + 2\mathcal{H}\delta' - c_s^2 \Delta \delta - \frac{\bar{\rho} a^2}{2M_{\text{pl}}^2} \delta = 0. \quad (1.123)$$

In Fourier space the equation is then written

$$\delta'' + 2\mathcal{H}\delta' + c_s^2 \left[ k^2 - \frac{\bar{\rho} a^2}{2c_s^2 M_{\text{pl}}^2} \right] \delta = 0. \quad (1.124)$$

Solutions depend on the sign of the last term. We introduce the comoving **Jeans wave-number**

$$k_{\text{J}} = \sqrt{\frac{\bar{\rho} a^2}{2c_s^2 M_{\text{pl}}^2}} = \frac{\sqrt{4\pi G_N \bar{\rho} a^2}}{c_s}, \quad (1.125)$$

and its associated length, the **Jeans length**  $\lambda_{\text{J}} = 2\pi a/k_{\text{J}}$ . This length defines the domains of the two possible solutions:

- **$k < k_{\text{J}}$** : There is a gravitational collapse, perturbations grow.
- **$k > k_{\text{J}}$** : The gravitational system is stable, the perturbations oscillate.

For the baryon component of large-scale structures, this Jeans length thus defines a scale for which the gas pressure prevents the gravitational collapse, effectively smoothing the gas density.

### Gravitational collapse

For scales larger than the Jeans length, the velocity dispersion is negligible and we can consider that  $\underline{\sigma} = 0$ . In this case, we recover the Meszaros 1.103 equation which corresponded in relativistic perturbation theory to the sub-Hubble case. One can then explicitly find the sub-Hubble equations with Newtonian gravity.

As a reminder, the two solutions of this equation in a matter dominated Universe ( $a \propto \eta^2$ ) are growing and decaying modes:

$$\begin{aligned} D_+(a) &\propto a, \\ D_-(a) &\propto a^{-3/2}. \end{aligned} \quad (1.126)$$

The  $D_+$  function, giving a gravitational instability, is called **linear growth function**. From this function, we define the **logarithmic growth rate of linear perturbations**  $f$  by

$$f = \frac{d \log D_+}{d \log a}. \quad (1.127)$$

This function depends on the cosmological model. It essentially scales with  $\Omega_m$ , so that it is usually parameterized by  $f = \Omega_m^\gamma$ . For the  $\Lambda$ CDM model,  $\gamma = 5/9$  is a good approximation, while deviations are expected in modified gravity models.

### The velocity field

It is possible to write the velocity as the sum of a part with no divergence and an irrotational part. Taking the curl of the velocity field (vorticity) in the second equation of 1.120, we obtain that the vorticity follows a law  $\nabla \times \vec{u} \propto a^{-1}$ . The expansion dilutes the vorticity and we can therefore neglect it and consider a solution with  $\nabla \times \vec{u} = \vec{0}$ . If we place ourselves in the case of a gravitational collapse, we keep the solution  $\delta \propto D_+$ . By noticing that in this case  $\delta' = \frac{D'_+}{D_+} \delta = f\mathcal{H}\delta$ , the continuity Eqn. 1.121 quickly gives

$$\frac{1}{\mathcal{H}} \nabla_x \cdot \vec{u} = -f\delta. \quad (1.128)$$

In Fourier space, the velocity components with no vorticity ( $\nabla \times \vec{u} = \vec{0}$ ) can be expressed as

$$\vec{u}(\vec{k}) = -i(\nabla_x \cdot \vec{u}) \frac{\vec{k}}{k^2}. \quad (1.129)$$

Using Eqn. 1.128, we simply obtain the following relation in Fourier space:

$$\vec{u}(\vec{k}) = if\mathcal{H} \frac{\vec{k}}{k^2} \delta(\vec{k}). \quad (1.130)$$

It is also useful to derive a general solution in configuration space. Eqn. 1.128 is similar to a Poisson-Boltzmann equation found in electrostatics. A solution in configuration space is

$$\vec{u}(\vec{x}) = \frac{f\mathcal{H}}{4\pi} \int d^3x' \delta(\vec{x}') \frac{\vec{x}' - \vec{x}}{|\vec{x}' - \vec{x}|^3}. \quad (1.131)$$

We will use a particular form of this solution to describe the dynamics of cosmic voids in Chap. 7 for which a spherical mass fluctuation is considered. It corresponds to the **spherical collapse** model. Considering a collapsing sphere of matter, let us define the comoving position  $\vec{x}_1$  of a fluid particle with respect to the center of this sphere. For the spherical geometry [51], Eqn. 1.131 becomes

$$\vec{u}(\vec{x}_1) = -\frac{f\mathcal{H}\vec{x}_1}{x_1^3} \int_0^{x_1} \delta(x'_1)x_1'^2 d^3x'_1 = -\frac{1}{3}f\mathcal{H}\bar{\delta}(x_1)\vec{x}_1, \quad (1.132)$$

where  $\bar{\delta}(x_1)$  is the average isotropic matter density contrast:

$$\bar{\delta}(x_1) = \frac{3}{x_1^3} \int_0^{x_1} \delta(x'_1)x_1'^2 d^3x'_1. \quad (1.133)$$

### Redshift space distortion

With spectroscopic surveys, it is possible to measure the logarithmic growth rate of linear perturbations  $f$  defined by Eqn. 1.127, using an observational effect due to peculiar velocities.

First of all, let us introduce the notion of bias. Observations of cosmic large-scale structures do not have direct access to the distribution of matter but only to one of its components (galaxies, quasars, gas...) because most of the matter is composed of dark matter. We say then that these tracers are biased tracers of the matter distribution. In the linear regime, it is possible to link the power spectrum of the density of a tracer  $t$  to that of the total matter by a **bias**  $b$ :

$$P_t(\vec{k}, z) = b^2(z)P_m(k, z). \quad (1.134)$$

Let us now take into account the effect of peculiar velocities on observations. Indeed, in a spectroscopic survey, the observation of a tracer is done in the space of redshifts. As indicated in Sec. 1.1.3, without strong gravitational lensing, the redshift is the sum of a component due to the expansion and the effect of particular velocities. Because of this, the relation between redshift and distance is not only due to expansion. Consequently, we speak about distance in the space of redshifts  $\vec{s}$  when we do not remove the effect of the particular velocities and only consider redshift. We speak of real space  $\vec{r}$  when we remove this effect. We will show how the peculiar velocities distort the clustering of a tracer  $t$  in redshift space.

The distance in redshift space  $\vec{s}_{ij}$  between two objects ( $i, j$ ) is related to the distance in real space  $\vec{r}_{ij}$  in the following way [52]:

$$\vec{s}_{ij} = \vec{r}_{ij} + (1+z) \frac{\hat{X}_{ij} \cdot (\vec{u}_i - \vec{u}_j)}{H(z)} \hat{X}_{ij}, \quad (1.135)$$

where  $\hat{X}_{ij}$  is the direction of observation towards the center of the object pair.

Considering a tracer  $t$  with respect to a fixed observer, and neglecting wide-angle effects, Eqn. 1.135 becomes

$$\vec{s} = \vec{r} + (1+z) \frac{u_r}{H(z)} \hat{r}, \quad (1.136)$$

where  $\hat{r} = \vec{r}/r$  gives the direction between the observer and the considered tracer, and  $u_r = \vec{u} \cdot \hat{r}$  is the velocity of the tracer along the  $\hat{r}$  direction.

We consider that the number of tracer is conserved through the transformation given by Eqn. 1.135. In this case, the contrast of tracer  $\delta$  expressed in real ( $\delta_t$ ) and redshift ( $\delta_t^s$ ) spaces verify:

$$(1 + \delta_t^s(\vec{s})) d^3\vec{s} = (1 + \delta_t(\vec{r})) d^3\vec{r}. \quad (1.137)$$

By expressing the Jacobian of Eqn. 1.136, the contrast in redshift space promptly becomes in Fourier space:

$$\delta_t^s(\vec{k}) = \int d^3\vec{r} \left[ \delta_t(\vec{r}) - (1+z) \frac{\partial_z u_r(\vec{r})}{H(z)} \right] e^{i(k\mu r/H + \vec{k}\cdot\vec{r})}, \quad (1.138)$$

where  $\mu$  is the cosine of the angle between  $\hat{r}$  and  $\vec{k}$ .

In the linear framework, we have seen that the peculiar velocities can be expressed as Eqn. 1.130. This link brings into play the logarithmic growth rate of linear perturbations  $f$ . By coupling this relation with Eqn. 1.138, a detailed calculation allows to obtain

$$\delta_t^s(\vec{k}) = (1 + \beta\mu^2)\delta_t(\vec{k}). \quad (1.139)$$

Finally, by adding this transformation in Eqn. 1.134, we obtain the **Kaiser formula** [52]:

$$P_t(z) = b^2(z)(1 + \beta\mu^2)^2 P_m(k, z), \quad (1.140)$$

where  $\beta = f/b$  is the **RSD parameter**.

The main result is the appearance of anisotropies in the power spectrum and simultaneously in the correlation function, which are called **redshift space distortions (RSD)**.

The impact of RSD will be considered in the context of the Lyman- $\alpha$  forest in Sec. 2.2.2 and in the context of its cross-correlation with voids in Chap. 7.

### Feedback from experiments

In practice, in Eqn. 1.140, the logarithmic growth rate of linear perturbations  $f$  is degenerate with the amplitude of the power spectrum. As a result, using appropriate RSD modeling for galaxy tracers, it is possible to obtain a measure of  $f\sigma_8$  (with  $\sigma_8$  defined by Eqn. 1.76). A major goal of DESI [26, 27] is to provide an accurate measurement of this parameter at different redshifts. By normalizing  $\sigma_8$  with other tracers, it is possible to deduce the value of  $f$ . An example of DESI forecast on the measurement of  $f$  is given in Fig. 1.14. This parameter is sensitive to variations with respect to general relativity. The measurement of this quantity is thus a powerful tool to test modified gravity models.

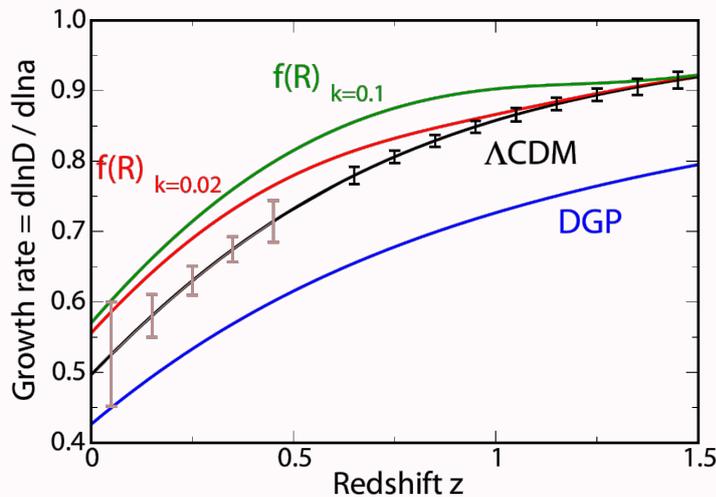


FIGURE 1.14 – Forecast of the measurement of  $f$  with the DESI instrument [26]. Some modified gravity models (f(R) and Dvali-Gabadadze-Porrati) are represented. The error bars obtained by DESI will be sufficient to discriminate gravity models.

### 1.3.3.2 Lagrangian approach

The Lagrangian approach consists in expressing the fluid properties not at their current place, but at their initial positions. By defining  $\vec{q}$ , the initial position of a fluid "particle", we define the current position of such a particle by

$$\vec{x} = \vec{q} + \vec{X}(\vec{q}, \eta). \quad (1.141)$$

It is then possible to relate the displacement field  $\vec{X}(\vec{q}, \eta)$  to the physical speed by

$$\vec{u} = \dot{\vec{x}} = \dot{\vec{X}}(\vec{q}, \eta). \quad (1.142)$$

In this Lagrangian case, the equation on the velocity does not involve a divergence term like  $\vec{u} \cdot \nabla_x \cdot \vec{u}$ :

$$\dot{\vec{u}} + \mathcal{H}\vec{u} = -\nabla_x \Phi. \quad (1.143)$$

By combining these two equations with the Poisson equation and applying the divergence, a comprehensive calculation yields the equation governing the evolution of the displacement field:

$$\nabla_x \cdot \ddot{\vec{X}} + 2\mathcal{H}\nabla_x \cdot \dot{\vec{X}} = \frac{1}{2}M_{\text{pl}}^{-2}a^2\bar{\rho}\delta. \quad (1.144)$$

It is possible to close the system of equations by expressing the density as a function of the Jacobian of the transformation  $\vec{q} \rightarrow \vec{x}$ . Non-linearities are then encoded by this transformation.

#### Zel'dovich

The **Zel'dovich approximation** [53] corresponds to a linear solution for the displacement field. It is obtained very similarly to the Eulerian case:

$$\nabla_q \cdot \ddot{\vec{X}} + 2\mathcal{H}\nabla_q \cdot \dot{\vec{X}} = \frac{3}{2}\Omega_m\mathcal{H}^2\nabla_q \cdot \vec{X}. \quad (1.145)$$

This equation is the equivalent of Meszaros Eqn. 1.103 in the Lagrangian approach. The Zel'dovich approximation consists in taking the growing solution of this equation such that

$$\vec{X}^{\text{Zel.}}(\vec{q}, t) = D_+(t)\vec{X}^+(\vec{q}), \quad (1.146)$$

where the solution  $D_+$  is kept in its general form as mentioned for Eqn. 1.103. The Lagrangian framework is very useful to define the initial conditions of a simulation. More generally, one can add terms of order 2 and higher in the Jacobian. This is called **n<sup>th</sup> order perturbation theory (nLPT)**. In particular, the 2LPT and 3LPT approximations are used in Chap. 3.

## 1.3.4 Impact of neutrino masses and warm dark matter

### 1.3.4.1 Neutrino masses

In Sec. 1.2.2.1, we presented the impact of neutrinos on the evolution of the background through the Friedmann Eqn. 1.15. In the case of massive neutrinos, this impact is mass-dependent as indicated by Eqn. 1.56.

Neutrino masses also influence perturbations. In principle, this impact is dependent on the number of relativistic or non-relativistic neutrino species. Therefore, its complete characterization depends on both the evolution of neutrinos in the Universe (interactions, expansion) and the full mass hierarchy of neutrinos detailed in Sec. 1.2.2.1. However, with approximation good enough for current and forthcoming cosmological observations, this impact depends primarily on  $\sum_{\nu} m_{\nu}$  whatever the number of non-relativistic species at  $z = 0$  and the hierarchy.

Neutrinos through their mass will modify both the CMB anisotropies and the formation of large-scale structures (LSS). For a complete description of the impact of neutrinos on the perturbation theory, references [33, 54, 55, 56] give more details.

#### Free-streaming

After their decoupling from the primordial plasma, neutrinos can be modeled by a collisionless fluid. Neutrinos then evolve as free particles. They are said to free-stream, with a thermal velocity  $v_{\text{th}}$ . It is possible to define the **free-streaming wavenumber** similarly to the Jeans wavenumber by replacing the sound speed  $c_s$  by the thermal speed. This quantity characterizes the comoving scale from which the free-streaming of neutrinos can be neglected in the evolution equations:

$$k_{\text{fs}}(t) = \sqrt{\frac{\bar{\rho}(t)a^2(t)}{2v_{\text{th}}^2(t)M_{\text{pl}}^2}} = \frac{\sqrt{4\pi G_N \bar{\rho}(t)a^2(t)}}{v_{\text{th}}(t)}. \quad (1.147)$$

The **free-streaming length**  $\lambda_{\text{fs}} = 2\pi a(t)/k_{\text{fs}}$  can be seen as a limit: neutrinos can not be confined inside a region smaller than this length. On the other hand, at a given time neutrinos can not travel over a distance higher than  $\lambda_{\text{fs}}$ , so that perturbations on scale  $k < k_{\text{fs}}$  are not affected by this effect.

When neutrinos are relativistic, their thermal velocity is  $c$  and the length  $\lambda_{\text{fs}}$  is the Hubble radius. When neutrinos are no longer relativistic, their thermal velocity is given by

$$v_{\text{th}} \equiv \frac{\langle p \rangle}{m} \simeq \frac{3T_{\nu}}{m}. \quad (1.148)$$

For neutrinos which become non-relativistic during the matter domination era, the free-streaming wavenumber reaches a minimum at the time of the transition to a non-relativistic state  $t_{\text{nr}}$  (with a corresponding scale factor  $a_{\text{nr}}$ ):

$$k_{\text{nr}} = k_{\text{fs}}(t_{\text{nr}}) \simeq 0.018 \Omega_{\text{m}}^{1/2} \left( \frac{m}{1 \text{ eV}} \right)^{1/2} h^{-1} \cdot \text{Mpc}. \quad (1.149)$$

Consequently, the  $k < k_{\text{nr}}$  modes are not affected by the mass of neutrinos.

### Impact on the CMB

Given our current knowledge, we know that  $f_\nu \leq 0.1$ : all neutrino species are relativistic before decoupling. Consequently, the impact of neutrino masses on the CMB is essentially an indirect effect. For a fixed amount of matter, increasing the neutrino density is equivalent to decreasing  $\Omega_{\text{cdm}}$ . The main effect on the CMB spectrum is through the modification of the time of equality:

$$a_{\text{eq}} = \frac{\omega_r}{\omega_b + \omega_{\text{cdm}}} = \frac{\omega_r}{\omega_m(1 - f_\nu)}, \quad (1.150)$$

with  $\omega_r$  including radiation (photons + massless neutrinos) [54]: the scale factor at equality depends on the neutrino abundance as  $a_{\text{eq}} \propto 1/(1 - f_\nu)$ . The result is an increase and a shift of the peak amplitude of the CMB power spectrum shown in Fig. 1.4.

### Impact on large-scale structures

In the context of this thesis, we focus more on the impact of neutrinos on the linear matter power spectrum. Considering the effect of neutrino free-streaming given by Eqn. 1.147, the impact on the perturbations and thus on the power spectrum will be scale-dependent.

To compute the matter power spectrum, it is not necessary to make a distinction between baryons and cold dark matter, except on very small scales. This is possible because after the recombination, the baryonic matter tends to follow the CDM and we have to a first approximation  $\delta_b = \delta_{\text{cdm}}$ . In the case of neutrinos, it will be necessary to differentiate the perturbations of each species. Indeed, due to their large velocities, neutrinos do not follow dark matter at all scales.

For scales  $k < k_{\text{nr}}$ , i.e. super-Hubble scales at the time of the non-relativistic transition, neutrino free-streaming did not have the time to damp their primordial fluctuations. In that case, neutrinos follow the baryons and the CDM.

For small scales  $k \gg k_{\text{nr}}$ , the power spectrum is impacted by neutrino free-streaming. As explained in [54], the impact of neutrinos is two fold:

- Neutrinos impact the power spectrum by their energetic density. Given their non-zero mass, they contribute to  $\rho_m$  and thus modify the Friedmann equation. And as their contribution to the perturbations  $\delta_m$  is very small because of free-streaming, this effect creates a diminution of the total linear matter power spectrum.
- The neutrino mass has a direct impact on the baryons and the CDM perturbations, as it modifies the perturbed gravitational potential.

We now consider one neutrino species which transitions to the non-relativistic regime at  $a_{\text{nr}} \gg a_{\text{eq}}$ , i.e. during the matter era, where we have  $\delta_b = \delta_{\text{cdm}}$ . In this case, the evolution of perturbations is given by Meszaros Eqn. 1.103. For  $k \gg k_{\text{nr}}$ , the neutrino perturbations are negligible compared to baryons and dark matter. Since  $\delta_b = \delta_{\text{cdm}}$ , we can replace  $\delta_m$  by  $\delta_{\text{cdm}}$  in Eqn. 1.103. In parallel, in the Hubble flow term, neutrinos contribute to the Friedmann equation:  $3\mathcal{H} = \frac{1}{2}M_{\text{pl}}^{-2}a^2(\bar{\rho}_{\text{cdm}} + \bar{\rho}_b + \bar{\rho}_\nu)$ . Consequently, the Meszaros Eqn. 1.105 is simply modified in the matter dominated era:

$$\delta_{\text{cdm}}'' + \frac{2}{\eta}\delta_{\text{cdm}}' - \frac{6}{\eta^2}(1 - f_\nu)\delta_{\text{cdm}} = 0. \quad (1.151)$$

The growing solution  $D_+$  is modified. For  $f_\nu \ll 1$ , we obtain:

$$\delta_{\text{cdm}} \propto a^{1-\frac{3}{5}f_\nu}. \quad (1.152)$$

Taking into account the shift of time of equality  $a_{\text{eq}}$  (Eqn. 1.150), which delays the growth of  $\delta_{\text{cdm}}$ , and also the effect of cosmological constant  $\Lambda$  through the  $g(a)$  function introduced in Eqn. 1.110, one obtains the following approximate relation [54]:

$$\frac{\delta_{\text{cdm}}^{\sum m_\nu}(a_0)}{\delta_{\text{cdm}}^{\sum m_\nu=0}(a_0)} = (1-f_\nu)^{1/2} \left( \frac{a_0 g(a_0)}{a_{\text{nr}}} \right)^{-\frac{3}{5}f_\nu}. \quad (1.153)$$

The total matter power spectrum, including neutrino perturbations, then writes:

$$\begin{aligned} P(k) &= \left\langle \left( \frac{\Omega_{\text{cdm}}\delta_{\text{cdm}} + \Omega_{\text{b}}\delta_{\text{b}} + \Omega_{\nu}\delta_{\nu}}{\Omega_{\text{cdm}} + \Omega_{\text{b}} + \Omega_{\nu}} \right)^2 \right\rangle \\ &= \begin{cases} \langle \delta_{\text{cdm}}^2 \rangle, & \text{for } k < k_{\text{nr}}, \\ [1-f_\nu]^2 \langle \delta_{\text{cdm}}^2 \rangle, & \text{for } k \gg k_{\text{nr}}. \end{cases} \end{aligned} \quad (1.154)$$

By combining Eqn. 1.154 and 1.153, we promptly obtain the ratio between the linear matter power spectra in the massive and massless case for scales  $k \gg k_{\text{nr}}$ :

$$\frac{P(k)^{\sum m_\nu}}{P(k)^{\sum m_\nu=0}} = (1-f_\nu)^3 \left( \frac{a_0 g(a_0)}{a_{\text{nr}}} \right)^{-\frac{6}{5}f_\nu} \simeq 1 - 8f_\nu. \quad (1.155)$$

The approximation obtained here is a very crude, historical formula, valid for  $f_\nu \ll 1$  and  $k \gg k_{\text{nr}}$ . For scales such that  $k \gg k_{\text{nr}}$ , the suppression of the power spectrum corresponds to a plateau.

### Calculation with a Boltzmann code

The impact of neutrino mass on the linear matter power spectrum can be computed precisely using Boltzmann codes. Considering a single non-relativistic neutrino species, the relative difference of the power spectrum for different masses is given in Fig. 1.15. On the other hand, Fig. 1.16 (left) gives an example of the dependence of the power spectrum on the number of massive neutrinos. In this example, the sum of masses is the same and is equally shared between different neutrino states. As shown in the figure, cosmological probes will be essentially sensitive to the sum of the neutrino masses, and not to the individual masses.

Finally, we show the impact of the redshift in Fig. 1.16 (right). There is some redshift dependence of the linear attenuation, degenerate with the effect of  $\sum_\nu m_\nu$ . Non-linear effects, not shown here, have a large impact on the predicted matter power spectrum.

To summarize, neutrinos impact by their mass the small scale clustering. We will see in Chap. 2 that the Lyman- $\alpha$  forest and in particular the measurement of its one-dimensional power spectrum detailed in Chap. 5, can exploit this effect to constrain the sum of neutrino masses.

#### 1.3.4.2 Warm dark matter

In the previous section, the impact of neutrinos as hot dark matter mixed with cold dark matter was considered. Current observations show that neutrinos cannot be the only component of the dark matter and the nature of dark matter is still unknown.

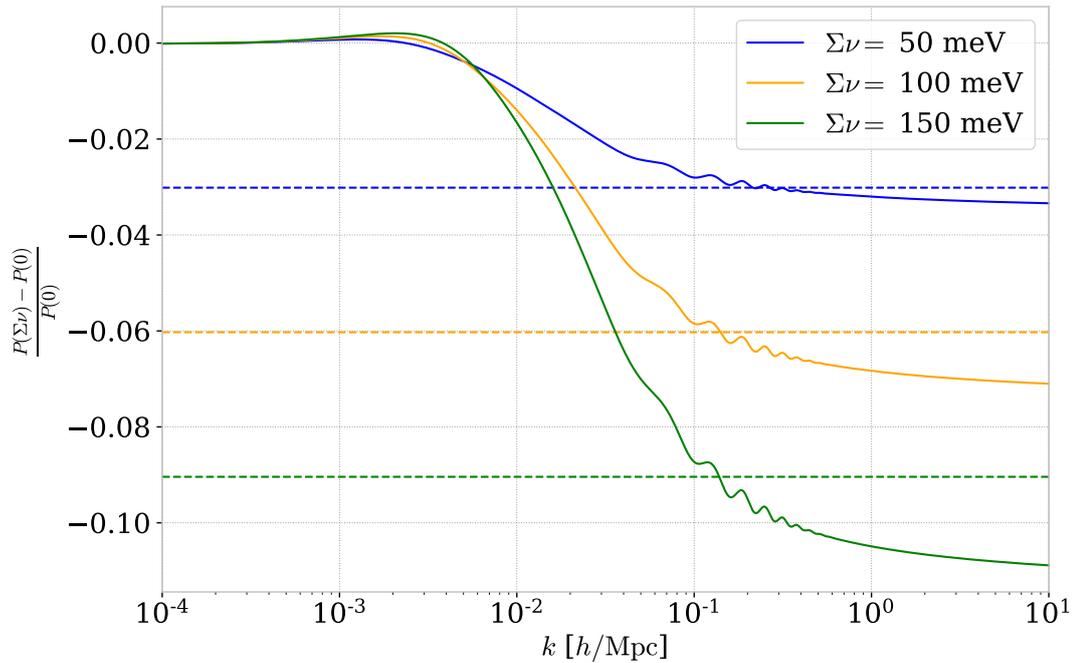


FIGURE 1.15 – Impact of the neutrino mass sum on the linear matter power spectrum. The quantity represented by continuous lines is the relative difference of the power spectrum between the massive and massless case. In the massive case, only one neutrino is considered massive. The dotted lines represent the analytical approximation  $-8f_\mu$  which works at  $k \gg k_{\text{nr}}$  and for low masses. These power spectra were obtained using the CLASS [15, 16] software with a flat  $\Lambda$ CDM model whose parameters are those of Planck 2018 [12].

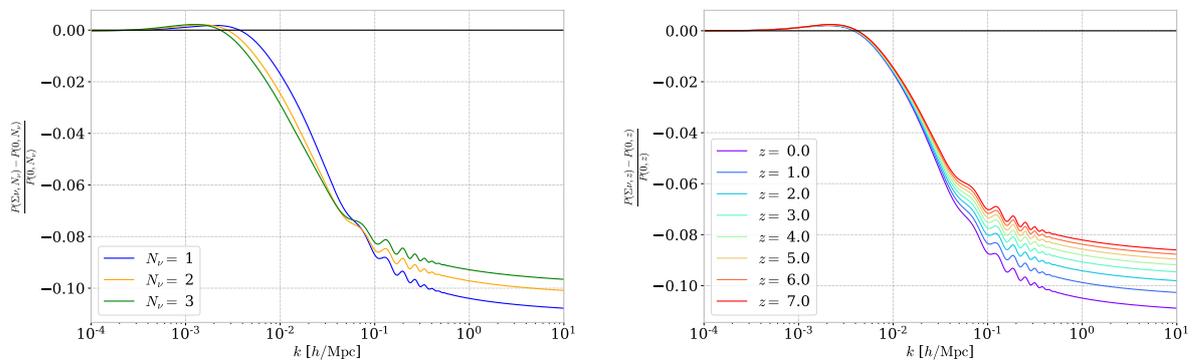


FIGURE 1.16 – (left) Relative difference of the linear matter power spectrum as a function of the number of neutrinos sharing a sum of masses of 150 meV. (right) Same quantity for one neutrino with a sum of masses of 150 meV expressed as a function of the redshift. These two plots have been obtained using the CLASS [15, 16] software with a flat  $\Lambda$ CDM model whose parameters are those of Planck 2018 [12].

A range of models called **warm dark matter (WDM)** has been considered both from particle physics and astrophysical considerations. A representation of the structure formation for different dark matter models is shown in figure 1.17. A model entirely composed of HDM would not be sufficient to explain the structures we observe today. The intermediate WDM case is still compatible with large-scale structure observations, provided the dark matter fluid is not "too" warm - which implies a lower bound in the keV range for the mass of a thermal relic [58].

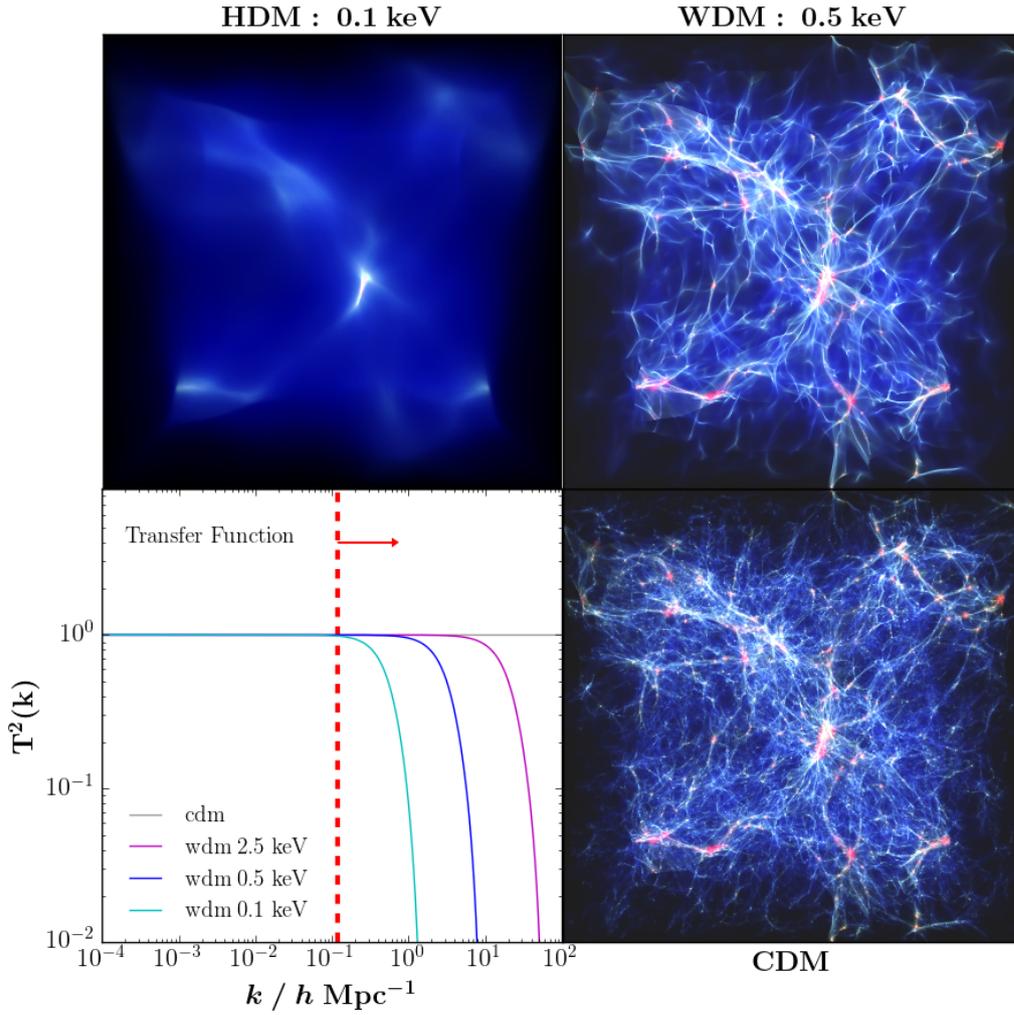


FIGURE 1.17 – (bottom left) Ratio between WDM and CDM linear power spectra considering different masses for the WDM case. (others) Volume rendering of the density field for simulations where the dark matter is Hot only, Warm only at 0.5 keV mass and Cold only. This plot was taken from [57].

Possible candidates for WDM can be found in theories beyond the standard model of particle physics. We will mention two types of candidates.

- **Thermal relics:** A thermal relic will make up WDM if it decouples from the primordial plasma at a time when it is still mildly relativistic, with a temperature  $T_x$ . It will then keep its Fermi momentum distribution (Eqn. 1.18 with a plus sign). Such a thermal relic contributes to change  $N_{\text{eff}}$  according to

$$\Delta N_{\text{eff}} = \left( \frac{T_x}{T_\nu} \right)^4. \quad (1.156)$$

Their non-relativistic energy density today can be related to their mass according to

$$\omega_x = \left( \frac{T_x}{T_\nu} \right)^3 \frac{m_x}{93.14\text{eV}} = (\Delta N_{\text{eff}})^{3/4} \frac{m_x}{93.14\text{eV}}. \quad (1.157)$$

- **Sterile neutrinos:** Sterile neutrinos are the hypothetical right-handed counterpart of ordinary (left-handed) neutrinos. They have no interactions with the W/Z bosons, but

they have some unknown mixing angle with active neutrinos, and unknown mass. The measurement of  $\Delta N_{\text{eff}}$  strongly constrains sterile neutrino scenarios as a thermal relic. However, other non-thermal models exist for the production of relic sterile neutrinos such as the case of non-resonantly produced sterile neutrinos. We will consider this model from now on. In this case, neutrinos have a re-scaled Fermi momentum distribution ( $\chi \ll 1$ ), related to their small mixing angle:

$$f_{\nu_s}(p) = \frac{\chi}{e^{p/T_\nu} + 1}. \quad (1.158)$$

Their impact on  $N_{\text{eff}}$  and their energy density can be related to their mass by the following equations:

$$\Delta N_{\text{eff}} = \chi \left( \frac{T_\nu}{\bar{T}_\nu} \right)^4 = \chi, \quad (1.159)$$

$$\omega_{\nu_s} = \chi \frac{m_{\nu_s}}{93.14 \text{eV}}. \quad (1.160)$$

### Impact on the power spectrum

As for neutrinos, one can define a free-streaming length depending on the thermal velocity of the considered WDM particle. If all of the dark matter is made of WDM, the impact on the linear power spectrum can be analytically fitted [59] to Boltzmann code predictions by

$$P_{\text{wdm}}(k, z) \simeq P_{\text{cdm}}(k, z) \left( 1 + \left( \frac{k}{k_0} \right)^{2\nu} \right)^{-\nu/5}, \quad (1.161)$$

where the parameter  $\nu = 1.12$  [60]. We can relate the pivotal scale  $k_0$  to the mass of the particle. When the mass of WDM particles is grown, free-streaming is restricted to smaller scales and the cut-off scale  $k_0$  grows accordingly.

In contrast to the  $\Lambda\text{CDM}\nu$  model, for modes  $k \gg k_0$ , the power spectrum is not only attenuated, it is fully suppressed. This comes from the fact that all of dark matter is considered warm, not like a mixed dark matter like the  $\Lambda\text{CDM}\nu$  model. This suppression is illustrated in Fig. 1.17 (bottom left).

In conclusion, WDM has a stronger impact on small scales than neutrinos mixed with CDM. The one dimensional power spectrum of the Lyman- $\alpha$  forest detailed in Chap. 5 is also a powerful tool to constrain WDM models.

## 1.4 Summary of cosmological observations

Throughout this chapter, examples of cosmological observations were given in boxes. The main message of these examples is that the construction of a viable model describing the Universe depends above all on observations.

The first type of probe encountered concerns the observation of the CMB. Some experiments like COBE/FIRAS measure spectral distortions of the CMB. A large source of cosmological constraints comes from the observation of CMB anisotropies presented in Sec. 1.2.1.2. Past observations such as WMAP [61] or Planck [12, 19] have in particular provided precise constraints on the  $\Lambda$ CDM model. Several ongoing experiments like ACT [62] or SPT [63] are already improving upon some of Planck's measurements. In the future, surveys like Simons Observatory [64] and LiteBird [65] will allow to improve even more the existing observations.

Another major class of cosmological observations is galaxy surveys. These observations include photometric galaxy surveys that observe the sky through a small number of integrated wavelength ranges called filters. A major goal of these surveys is to detect images of galaxies distorted by foreground material. For those, there are two types of lensing regimes: the strong lensing (SL) regime where the image is strongly distorted by compact or very massive objects. The weak lensing (WL) regime corresponds to the weak deformation generated by the cosmic web on a photometric image. Current surveys like DES [49] allow to constrain the  $\Lambda$ CDM model by the determination of the matter distribution. Future missions on this subject like Euclid [66, 67] or the Rubin Observatory [68] will improve over sky coverage and depth.

Another type of galaxy survey called spectroscopic focuses on the observation of the spectra of individual sources. They allow to determine precisely the redshift of the observed objects and to draw up a catalog. These catalogs are then used to measure clustering properties, giving access to the measurement of the BAO peak, of RSD, and of the distortions of the correlation functions due to the choice of a bad fiducial cosmology (Alcock-Paczyński [69] test). The obtained spectra also have many more information than their source's redshifts. In particular, the observation of quasars with a redshift higher than  $z \simeq 2$  allows to measure the Lyman- $\alpha$  forest. This is the main subject of this thesis presented in Chap. 2. Past surveys such as eBOSS [25] and ongoing ones such as DESI [26, 27] allow precise tests of the  $\Lambda$ CDM model and of its possible extensions. Future experiments like the spectrograph part of Euclid [66, 67] or WEAVE-QSO [70] will also improve these measurements.

Radio astronomy is also an interesting field for cosmology with the detection of the 21 cm line. It corresponds to the transition between the two hyper-fine levels of the ground state of neutral hydrogen. The signal at 21 cm remains weak and hardly discernible compared to all other contaminating foregrounds. This detection has strong potential because it could map the distribution of neutral hydrogen in the Universe, over a wide range in redshift from the dark ages to today ( $100 < z < 0$ ). The Canadian Hydrogen Intensity Mapping Experiment (CHIME) [72] has already detected a cross-correlation signal with SDSS [73]. Future surveys carried out with the Square Kilometre Array (SKA) [74] will surely bring very valuable results for cosmology.

Finally, the last major class of cosmological observables is the detection of Type Ia supernovae (SNIa) which was the first observation demonstrating the accelerated expansion of the Universe. Recent surveys such as ZTF [75] and future ones such as the Rubin Observatory [68] will provide a large catalog of supernovae.

There are many other cosmological observables. The probes listed above are very complementary. Combining several types of observations, or probes, has now become a powerful tool to

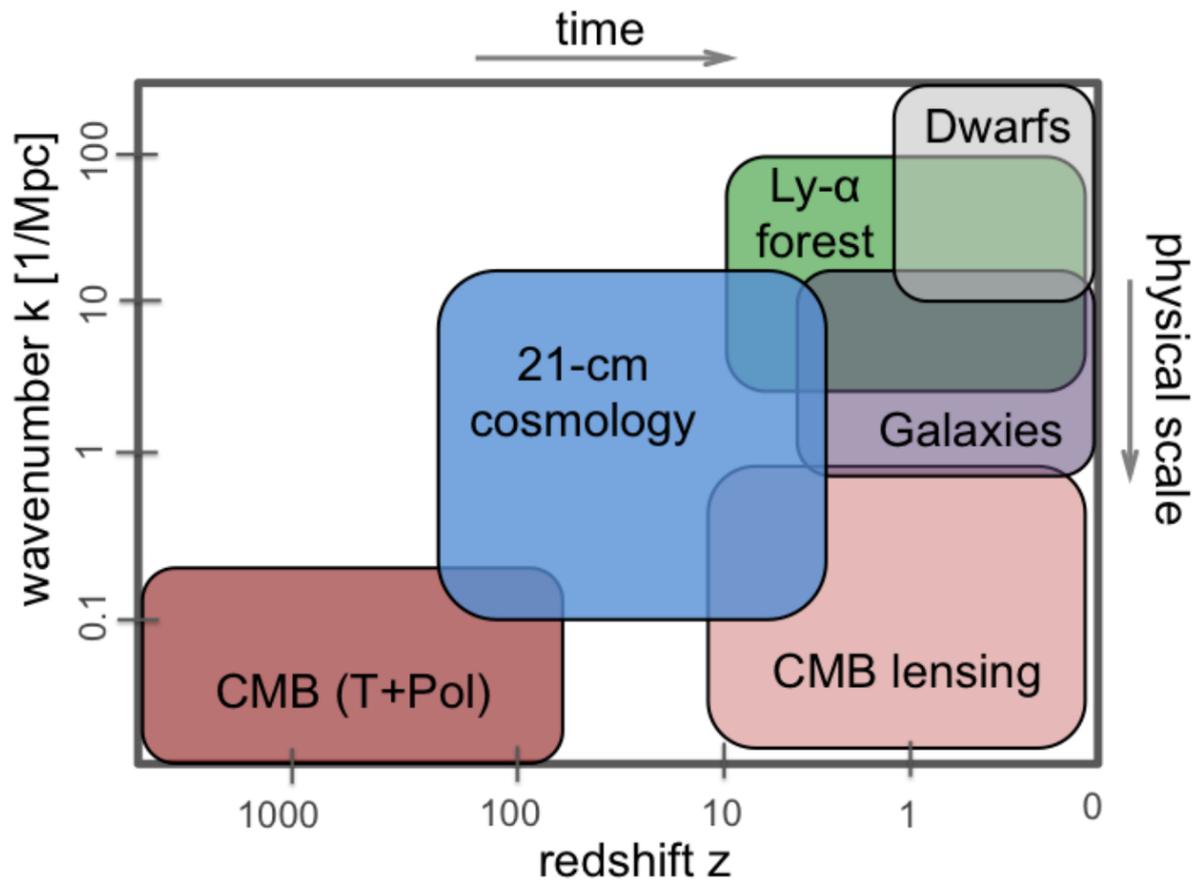


FIGURE 1.18 – Schematic representation of the approximate scales and epochs probed by different cosmological large-scale structure observations, taken from [71].

mitigate the impact of systematics related to each type of probes, while improving cosmological constraints by exploiting different degeneracies on cosmological parameters. The tracers listed above probe very different redshift and spatial scales as illustrated in Fig. 1.18.

## Bibliography

- [1] P. Peter and J.-P. Uzan, *Primordial cosmology*. Oxford University Press, Oxford ; New York, english ed ed., 2009.
- [2] F. Bernardeau, *Cosmologie: Des fondements théoriques aux observations*. EDP Sciences, 2007, [10.1051/978-2-7598-0101-5](#).
- [3] S. Dodelson, *Modern cosmology*. Academic Press, San Diego, Calif, 2003.
- [4] J. Rich, *Fundamentals of Cosmology*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, [10.1007/978-3-642-02800-7](#).
- [5] J. Baur, *Determining the mass of cosmological neutrinos using Lyman-alpha forests*, phdthesis, Université Paris Saclay (COmUE), Sept., 2017.
- [6] A. D. Mattia, *Des relevés spectroscopiques de galaxies à raies d'émission aux contraintes cosmologiques : d'eBOSS à DESI*, phdthesis, Université Paris-Saclay, Sept., 2020.
- [7] S. Chabanier, *Neutrinos and dark matter cosmology with the Lyman- $\alpha$  forest : the interplay between large-scale evolution and small-scale baryonic physics*, phdthesis, Université Paris-Saclay, Oct., 2020.
- [8] D. W. Hogg, *Distance measures in cosmology*, *arXiv:astro-ph/9905116* (2000) .
- [9] G. Lemaître, *Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques*, *Annales de la Société Scientifique de Bruxelles* **47** (1927) 49.
- [10] E. Hubble, *A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae*, *Proceedings of the National Academy of Science* **15** (1929) 168.
- [11] A. G. Riess, S. Casertano, W. Yuan, J. B. Bowers, L. Macri, J. C. Zinn et al., *Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with  $\Lambda$ CDM*, *The Astrophysical Journal Letters* **908** (2021) L6.
- [12] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astronomy & Astrophysics* **641** (2020) A6.
- [13] W. L. Freedman, *Measurements of the Hubble Constant: Tensions in Perspective*, *The Astrophysical Journal* **919** (2021) 16.
- [14] P. Shah, P. Lemos and O. Lahav, *A buyer's guide to the Hubble Constant*, *arXiv:2109.01161 [astro-ph]* (2021) .
- [15] J. Lesgourgues, *The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview*, *arXiv:1104.2932 [astro-ph]* (2011) .
- [16] D. Blas, J. Lesgourgues and T. Tram, *The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes*, *Journal of Cosmology and Astroparticle Physics* **2011** (2011) 034.

- [17] M. Fumagalli, J. M. O’Meara and J. X. Prochaska, *Detection of Pristine Gas Two Billion Years After the Big Bang*, *Science* **334** (2011) 1245.
- [18] R. Cooke, M. Pettini and C. C. Steidel, *One percent determination of the primordial deuterium abundance*, *The Astrophysical Journal* **855** (2018) 102.
- [19] PLANCK collaboration, *Planck 2018 results. I. Overview and the cosmological legacy of Planck*, *Astronomy & Astrophysics* **641** (2020) A1.
- [20] K. S. Dawson, J.-P. Kneib, W. J. Percival, S. Alam, F. D. Albareti, S. F. Anderson et al., *The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data*, *The Astronomical Journal* **151** (2016) 44.
- [21] M. R. Blanton, M. A. Bershady, B. Abolfathi, F. D. Albareti, C. A. Prieto, A. Almeida et al., *Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe*, *arXiv:1703.00052 [astro-ph]* (2017) .
- [22] S. A. Smee, J. E. Gunn, A. Uomoto, N. Roe, D. Schlegel, C. M. Rockosi et al., *The Multi-object, Fiber-fed Spectrographs for the Sloan Digital Sky Survey and the Baryon Oscillation Spectroscopic Survey*, *The Astronomical Journal* **146** (2013) 32.
- [23] J. E. Gunn, W. A. Siegmund, E. J. Mannery, R. E. Owen, C. L. Hull, R. F. Leger et al., *The 2.5 m Telescope of the Sloan Digital Sky Survey*, *The Astronomical Journal* **131** (2006) 2332.
- [24] D. G. York, *The Sloan Digital Sky Survey: Technical Summary*, *The Astronomical Journal* **120** (2000) 1579.
- [25] R. Ahumada, C. A. Prieto, A. Almeida, F. Anders, S. F. Anderson, B. H. Andrews et al., *The Sixteenth Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra*, *arXiv:1912.02905 [astro-ph]* (2020) .
- [26] DESI collaboration, *The DESI Experiment Part I: Science, Targeting, and Survey Design*, *arXiv:1611.00036 [astro-ph]* (2016) .
- [27] DESI collaboration, *The DESI Experiment Part II: Instrument Design*, *arXiv:1611.00037 [astro-ph]* (2016) .
- [28] R. H. Becker, X. Fan, R. L. White, M. A. Strauss, V. K. Narayanan, R. H. Lupton et al., *Evidence for Reionization at  $z \sim 6$ : Detection of a Gunn-Peterson Trough in a  $z=6.28$  Quasar*, *The Astronomical Journal* **122** (2001) 2850.
- [29] M. Dijkstra, *Physics of Lyman Alpha Radiative Transfer*, *arXiv:1704.03416 [astro-ph]* (2017) .
- [30] D. J. Fixsen, *The Temperature of the Cosmic Microwave Background*, *The Astrophysical Journal* **707** (2009) 916.
- [31] F. R. C. L. Cowan, F. B. Harrison, H. W. Kruse and A. D. McGuire, *Detection of the Free Neutrino: A Confirmation*, *Science, New Series* **124** (1956) 103.

- [32] SUPER-KAMIOKANDE collaboration, *Evidence for Oscillation of Atmospheric Neutrinos*, *Physical Review Letters* **81** (1998) 1562.
- [33] J. Lesgourgues, G. Mangano, G. Miele and S. Pastor, *Neutrino Cosmology*. Cambridge University Press, Feb., 2013.
- [34] P. F. de Salas, S. Gariazzo, O. Mena, C. A. Ternes and M. Tórtola, *Neutrino Mass Ordering from Oscillations and Beyond: 2018 Status and Future Prospects*, *Frontiers in Astronomy and Space Sciences* **5** (2018) 36.
- [35] THE CUPID INTEREST GROUP collaboration, *CUPID pre-CDR*, *arXiv:1907.09376* (2019) .
- [36] J. Wolf, *The KATRIN Neutrino Mass Experiment*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **623** (2010) 442.
- [37] M. Aker, A. Beglarian, J. Behrens, A. Berlev, U. Besserer, B. Bieringer et al., *First direct neutrino-mass measurement with sub-eV sensitivity*, *arXiv:2105.08533 [hep-ex, physics:nucl-ex]* (2021) .
- [38] EBOSS collaboration, *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological Implications from two Decades of Spectroscopic Surveys at the Apache Point observatory*, *Physical Review D* **103** (2021) 083533.
- [39] K. G. Begeman, A. H. Broeils and R. H. Sanders, *Extended rotation curves of spiral galaxies: dark haloes and modified dynamics*, *Monthly Notices of the Royal Astronomical Society* **249** (1991) 523.
- [40] M. Markevitch, A. H. Gonzalez, D. Clowe, A. Vikhlinin, L. David, W. Forman et al., *Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56*, *The Astrophysical Journal* **606** (2004) 819.
- [41] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiattia, A. Diercks, P. M. Garnavich et al., *Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant*, *The Astronomical Journal* **116** (1998) 1009.
- [42] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro et al., *Measurements of  $\Omega$  and  $\Lambda$  from 42 High-Redshift Supernovae*, *The Astrophysical Journal* **517** (1999) 565.
- [43] M. Betoule, R. Kessler, J. Guy, J. Mosher, D. Hardin, R. Biswas et al., *Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples*, *Astronomy & Astrophysics* **568** (2014) A22.
- [44] M. Chevallier and D. Polarski, *Accelerating Universes with Scaling Dark Matter*, *International Journal of Modern Physics D* **10** (2001) 213.
- [45] E. V. Linder, *Exploring the Expansion History of the Universe*, *Physical Review Letters* **90** (2003) 091301.
- [46] S. Weinberg, *Cosmological Fluctuations of Short Wavelength*, *The Astrophysical Journal* **581** (2002) 810.

- [47] P. Meszaros, *The behaviour of point masses in an expanding cosmological substratum.*, *Astronomy and Astrophysics* **37** (1974) 225.
- [48] A. Lewis, A. Challinor and A. Lasenby, *Efficient Computation of CMB anisotropies in closed FRW models*, *The Astrophysical Journal* **538** (2000) 473.
- [49] M. Gatti, E. Sheldon, A. Amon, M. Becker, M. Troxel, A. Choi et al., *Dark Energy Survey Year 3 Results: Weak Lensing Shape Catalogue*, *Monthly Notices of the Royal Astronomical Society* **504** (2021) 4312.
- [50] S. Chabanier, M. Millea and N. Palanque-Delabrouille, *Matter power spectrum: from Ly $\alpha$  forest to CMB scales*, *Monthly Notices of the Royal Astronomical Society* **489** (2019) 2247.
- [51] P. J. E. Peebles, *Principles of physical cosmology*. princeton university press ed., 1993.
- [52] N. Kaiser, *Clustering in real space and in redshift space*, *Monthly Notices of the Royal Astronomical Society* **227** (1987) 1.
- [53] Y. Zel'dovich, *Gravitational instability: An Approximate theory for large density perturbations*, *undefined* (1969) .
- [54] J. Lesgourgues and S. Pastor, *Massive neutrinos and cosmology*, *Physics Reports* **429** (2006) 307.
- [55] J. Lesgourgues and S. Pastor, *Neutrino Mass from Cosmology*, *Advances in High Energy Physics* **2012** (2012) 1.
- [56] J. Lesgourgues and S. Pastor, *Neutrino cosmology and Planck*, *New Journal of Physics* **16** (2014) 065002.
- [57] J. Baur, N. Palanque-Delabrouille, C. Yèche, C. Magneville and M. Viel, *Lyman-alpha Forests cool Warm Dark Matter*, *Journal of Cosmology and Astroparticle Physics* **2016** (2016) 012.
- [58] S. Tremaine and J. E. Gunn, *Dynamical Role of Light Neutral Leptons in Cosmology*, *Physical Review Letters* **42** (1979) 407.
- [59] P. Bode, J. P. Ostriker and N. Turok, *Halo Formation in Warm Dark Matter Models*, *The Astrophysical Journal* **556** (2001) 93.
- [60] M. Viel, J. Lesgourgues, M. G. Haehnelt, S. Matarrese and A. Riotto, *Constraining Warm Dark Matter candidates including sterile neutrinos and light gravitinos with WMAP and the Lyman-alpha forest*, *Physical Review D* **71** (2005) 063534.
- [61] G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley et al., *Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results*, *The Astrophysical Journal Supplement Series* **208** (2013) 19.
- [62] A. D. Hincks, V. Acquaviva, P. A. R. Ade, P. Aguirre, M. Amiri, J. W. Appel et al., *The atacama cosmology telescope (ACT): Beam profiles and first SZ cluster maps*, *The Astrophysical Journal Supplement Series* **191** (2010) 423.

- [63] J. E. Carlstrom, P. A. R. Ade, K. A. Aird, B. A. Benson, L. E. Bleem, S. Buseti et al., *The 10 Meter South Pole Telescope*, *Publications of the Astronomical Society of the Pacific* **123** (2011) 568.
- [64] THE SIMONS OBSERVATORY collaboration, *The Simons Observatory: Astro2020 Decadal Project Whitepaper*, *arXiv:1907.08284 [astro-ph]* (2019) .
- [65] M. Hazumi, P. A. R. Ade, A. Adler, E. Allys, K. Arnold, D. Auguste et al., *LiteBIRD: JAXA's new strategic L-class mission for all-sky surveys of cosmic microwave background polarization*, *Space Telescopes and Instrumentation 2020: Optical, Infrared, and Millimeter Wave* (2020) 249.
- [66] R. Laureijs, J. Amiaux, S. Arduini, J.-L. Auguères, J. Brinchmann, R. Cole et al., *Euclid Definition Study Report*, *arXiv:1110.3193 [astro-ph]* (2011) .
- [67] G. D. Racca, R. Laureijs, L. Stagnaro, J. C. Salvignol, J. L. Alvarez, G. S. Criado et al., *The Euclid mission design*, *arXiv:1610.05508 [astro-ph]* (2016) 990400.
- [68] Ž. Ivezić, S. M. Kahn, J. A. Tyson, B. Abel, E. Acosta, R. Allsman et al., *LSST: From Science Drivers to Reference Design and Anticipated Data Products*, *The Astrophysical Journal* **873** (2019) 111.
- [69] C. Alcock and B. Paczyński, *An evolution free test for non-zero cosmological constant*, *Nature* **281** (1979) 358.
- [70] M. M. Pieri, S. Bonoli, J. Chaves-Montero, I. Paris, M. Fumagalli, J. S. Bolton et al., *WEAVE-QSO: A Massive Intergalactic Medium Survey for the William Herschel Telescope*, *arXiv:1611.09388 [astro-ph]* (2016) .
- [71] V. Gluscevic, Y. Ali-Haimoud, K. Bechtol, K. K. Boddy, C. Boehm, J. Chluba et al., *Cosmological Probes of Dark Matter Interactions: The Next Decade*, *arXiv:1903.05140 [astro-ph]* (2019) .
- [72] a. M. Amiri, K. Bandura, P. Berger, M. Bhardwaj, M. M. Boyce, P. J. Boyle et al., *The CHIME Fast Radio Burst Project: System Overview*, *The Astrophysical Journal* **863** (2018) 48.
- [73] CHIME collaboration, *Detection of Cosmological 21 cm Emission with the Canadian Hydrogen Intensity Mapping Experiment*, *arXiv:2202.01242 [astro-ph]* (2022) .
- [74] P. Dewdney, P. Hall, R. Schilizzi and T. Lazio, *The Square Kilometre Array*, *Proceedings of the IEEE* **97** (2009) 1482.
- [75] E. C. Bellm, S. R. Kulkarni, M. J. Graham, R. Dekany, R. M. Smith, R. Riddle et al., *The Zwicky Transient Facility: System Overview, Performance, and First Results*, *Publications of the Astronomical Society of the Pacific* **131** (2018) 018002.

# 2

---

## The Lyman- $\alpha$ forest

---

*“Activity does not necessarily mean life. Quasars are active. And a monk meditating is not inanimate.”*

– Philip K. Dick, *A Scanner Darkly*, 1977

**Contents**

---

<b>2.1</b>	<b>Properties of the Lyman-<math>\alpha</math> forest</b>	<b>62</b>
2.1.1	Quasars	62
2.1.2	Lyman- $\alpha$ absorption	65
2.1.2.1	The intergalactic medium	65
2.1.2.2	Lyman- $\alpha$ forest structure	66
2.1.3	Astrophysical systematics of Lyman- $\alpha$ forest	71
<b>2.2</b>	<b>Clustering with the Lyman-<math>\alpha</math> forest</b>	<b>74</b>
2.2.1	Lyman- $\alpha$ forest correlations	74
2.2.2	Modeling of bias and redshift space distortions	75
<b>2.3</b>	<b>Some applications of the Lyman-<math>\alpha</math> forest in cosmology</b>	<b>79</b>
2.3.1	Observation types	79
2.3.2	Lyman- $\alpha$ forest based measurements	81
	<b>Bibliography</b>	<b>86</b>

---

---

AMONG all the cosmological tracers presented in the previous chapter, this thesis uses the Lyman- $\alpha$  (hereafter Ly $\alpha$ ) forest. Compared to point tracers, such as galaxies, the Ly $\alpha$  forest stands out by its ability to probe small-scale fluctuations of the cosmic web. It is thus a powerful cosmological probe allowing very different analyses with respect to other tracers of large-scale structure surveys.

The Ly $\alpha$  forest can be observed by measuring the spectrum of quasars, which are among the brightest astrophysical objects known to date. The nature of these objects has long remained mysterious. Thanks to advances in astrophysics, it is now known that the intense light of quasars comes from fast-rotating matter around a supermassive black hole. The emitted light goes through the cosmic web located between the quasar and the observer. On their way, photons are partly absorbed by elements of the intergalactic medium. In particular, the Ly $\alpha$  absorptions caused by neutral hydrogen located at different redshifts imprint a sequence of absorption lines on the quasar spectrum. These features constitute the Ly $\alpha$  forest. A description of quasars and the formation of the Ly $\alpha$  forest is given in Sec. 2.1.

The Ly $\alpha$  forest is thus a tracer of neutral hydrogen in the intergalactic medium. The best way to extract cosmological information from this signal is to study the correlations related to the Ly $\alpha$  forest. I describe associated observables in Sec. 2.2.

In modern cosmology, the Ly $\alpha$  forest is the object of growing interest. Many Ly $\alpha$  studies have already yielded significant cosmological constraints. In Sec. 2.3, some surveys and analyses using the Ly $\alpha$  forest will be detailed. In particular, I will highlight how my thesis is embedded in this context.

This chapter is based on the reference books [1, 2, 3, 4], on the courses of "galaxies et cosmologie" of Françoise Combes at "Collège de France" and on previous manuscripts [5, 6, 7, 8].

## 2.1 Properties of the Lyman- $\alpha$ forest

### 2.1.1 Quasars

In 1939, the amateur radio astronomer Grote Reber discovered peculiar extragalactic objects emitting in the optical and radio ranges without understanding their nature. During the year 1943, Carl Seyfert discovered galaxies with very bright nuclei. Their spectra display a continuum, very broad emission lines, and some absorptions. They emit in the optical range and sometimes in radio. These first galaxies allowed to explain part of the objects found by Reber but not in totality.

In 1963, Thomas Matthews and Allan Sandage [9] detected sources which present the same properties as the Seyfert galaxies, but that are point-like. The spectra of these objects were then measured by Jesse Greenstein and Maarten Schmidt [10]. In particular, one of these objects is observed at  $z = 0.158$ , an extremely large redshift at the time. This source is named 3C273 and constitutes the first observation of a quasar. Since this observation, many extragalactic objects with similar properties have been observed.

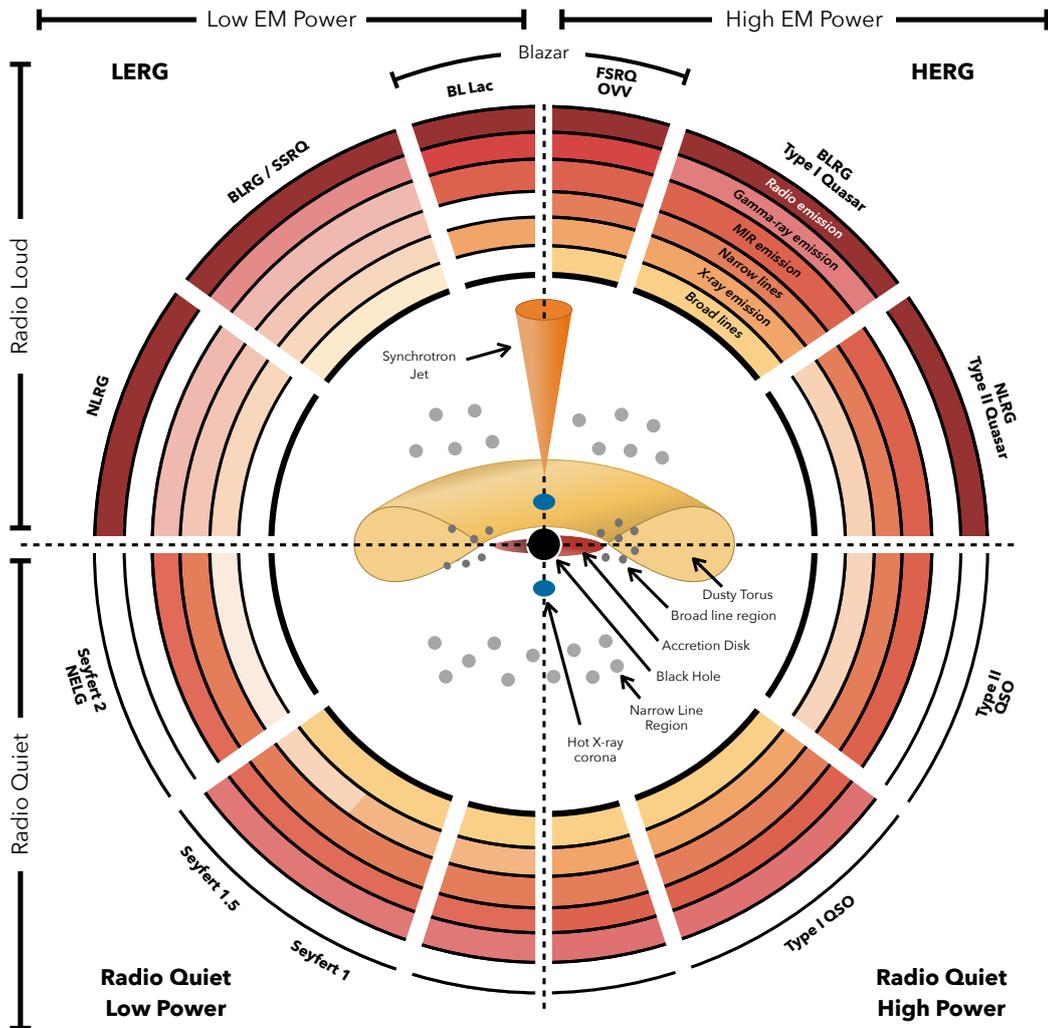
#### Active Galactic Nuclei

All these objects have been categorized as **active galactic nuclei (AGN)**. Indeed, the explanation that best fits the observed spectra is the existence of a zone of high activity at the heart of a galaxy, sourced by an extremely large mass in its center.

Active galactic nuclei are classified into several subclasses. Their common characteristics are the presence of a very high continuum spectrum, the containment to a small volume, and the variability of their emission. Their differentiation is made from the presence of radio emission, their brightness, and the importance of their variability. The six main AGN categories are:

- **Seyfert Galaxies:** Spatially extended galaxies which present a spectrum with narrow and sometimes broad lines. They are called Seyfert 1 when they have a lot of broad emission lines and Seyfert 2 when this is not the case. They show strong emission in the visible and X-ray range, and sometimes a little in the radio range. Their luminosity is approximately 100 times greater than the Milky Way.
- **Quasars:** Sources very similar to the Seyfert 1 galaxies but which do not show spatial extension. Their luminosity is more significant and can reach 1000 times the Milky Way. They can be radio-loud or radio-quiet.
- **Low-ionization nuclear emission-line regions (LINERs):** The least luminous galaxies among AGN, which are characterized by emission lines indicating the presence of weakly ionized elements.
- **Blazars:** Highly variable sources emitting strongly in the X-ray and radio range but showing no emission lines.
- **Radio galaxies:** This group comprises all galaxies with a fainter luminosity than Seyfert ones, and emitting in the radio domain.

The main disadvantage of this classification into subcategories is that it is related to the history of observations rather than the actual physical processes. A unification model was developed to understand the AGN diversity in a common picture. The first premises of this unification was the creation of a model coupling a supermassive object at the center of the AGN with a disk of



J. E. Thorne

FIGURE 2.1 – Representation of all possible types of AGN on a unification diagram taken from [11]. The nature of an AGN depends on its electromagnetic radiation power, the presence of a relativistic jet, the composition of its environment, and its orientation with respect to the observer. The energy emitted in the form of radiation ultimately comes from the gravitational energy of the central supermassive black hole.

matter (see e.g. [12, 13]). This hypothesis allowed to explain the Seyfert galaxies of type 1 and 2 simultaneously. It was confirmed observationally in [14].

The best model to date is the AGN unification model [15]. It helps explaining the diversity of AGN by modelling them as a system composed of a supermassive black hole ( $10^6$  to  $10^{10}$  solar masses), an accretion disk, high- and low- velocity gas, an obscuring torus, and a relativistic jet. A graphical representation of this model is given in Fig. 2.1. All of the AGN can be explained by the variation of the physical properties of each of these elements. In addition, the diversity of the observed spectra is explained by the orientation of the AGN with respect to the observer.

### Quasar properties

This thesis focuses on the observation of **quasars** (contraction of quasi-stars), also called **quasi-stellar objects (QSOs)**.

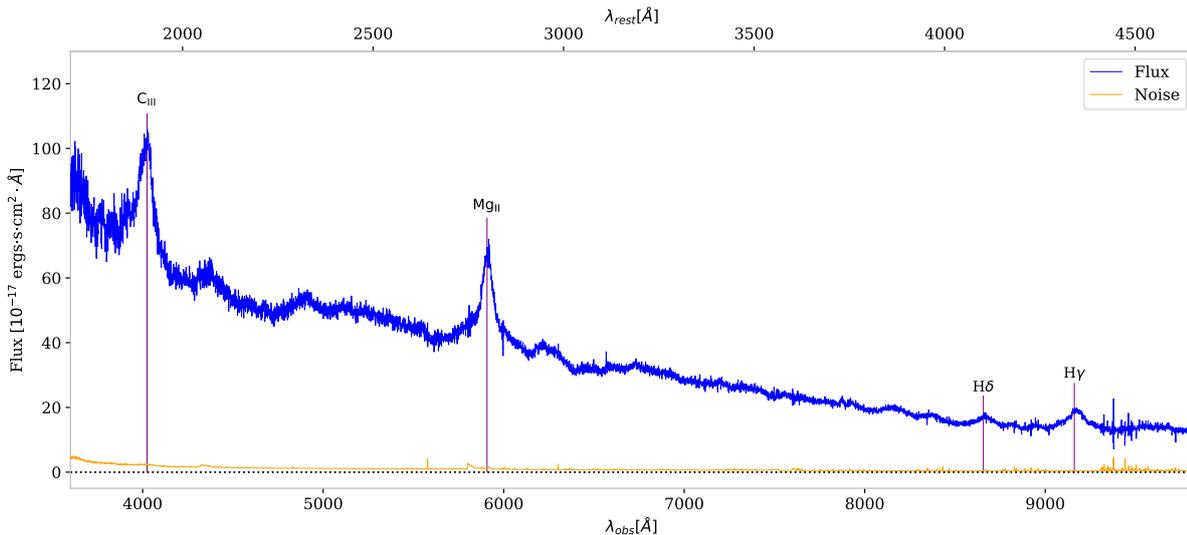


FIGURE 2.2 – The spectrum of a quasar located at redshift  $z = 1.11$  measured by DESI with an exposure time of 1900 seconds. It is a very bright quasar whose flux is represented in blue, and the instrumental noise in orange. The quasar spectrum is characterized by a strong continuum and broad emission lines.

More precisely, the objects used in this thesis correspond mostly to type 1 quasar in the AGN unification model. They are thus characterized by a strong emission in the visible, X-ray, and gamma-ray domains with broad emission lines. They can be radio-quiet or radio-loud and present a weak infrared emission due to the dusty torus. Their enormous luminosity arises from gravitational energy and makes them the brightest objects in the Universe. The QSO core is concentrated in a volume of radius about 10 parsecs. It is composed of a central supermassive black hole that accretes the surrounding matter. The viscosity of the accretion disk induces thermal and non-thermal processes that cause the heating of the matter. This matter then emits a light responsible for the continuous emission of the QSO called **quasar continuum** and the emission lines. An example of a QSO spectrum measured by DESI is given in Fig. 2.2.

The emission lines have a width ranging from 1,000 to 40,000  $\text{km}\cdot\text{s}^{-1}$ . This large width is associated with the heated gas very close to the quasar. A quasar spectrum can also display some narrow lines (between 200 and 1,000  $\text{km}\cdot\text{s}^{-1}$ ) due to the heating of the more distant gas. The main elements responsible for the emission lines in the QSO spectrum are hydrogen, carbon, and magnesium. Some lines are more minorly produced by oxygen, silicon, nitrogen, and iron. The ionization level of these elements ranges from one- to five-times ionized for oxygen and even higher for iron.

Quasars generally exhibit luminosity variations on a very wide range of time scales, from the day to more than ten years. This variability is advantageous to help distinguish QSOs from stars which are the main contaminants in a photometric selection.

The farthest QSO currently detected is located at redshift  $z = 7.642$  [16]. Quasars used in this thesis have a redshift between  $z = 2.2$  and  $z = 5.0$ . Thus, they are seen as they were more than 10 billion years ago.

In my thesis, I used quasar spectra to measure the Ly $\alpha$  forest. However, quasars themselves are tracers of matter since they are located preferentially in the high-density regions of the Universe. When we consider QSOs as tracers, it is possible to relate their density contrast to that of matter on large scales:

$$\delta_{\text{QSO}} = b_{\text{q}} \delta_{\text{m}}. \quad (2.1)$$

Depending on the model considered, the **quasar bias**  $b_{\text{q}}$  can be linear or more complex. It depends on the redshift. Due to their rarity, quasars are highly-biased tracers. The typical value of their bias is 2.5 for  $z \sim 1.5$  [17].

## 2.1.2 Lyman- $\alpha$ absorption

Quasar spectra also contain absorption lines. Except for the case of broad absorption line (BAL) QSOs described in Sec. 2.1.3, these lines are due to matter located outside the quasar, along the line-of-sight. This is notably the case for the **Ly $\alpha$  forest** which is the main object of study of this thesis and which probes neutral hydrogen present between galaxies.

### 2.1.2.1 The intergalactic medium

The **intergalactic medium (IGM)** is a very diffuse plasma located between galaxies. As it is only fed by very small amounts of stellar nucleosynthesis products, its composition is close to that of the primordial plasma of the Universe, as determined by BBN (see Sec. 1.2.1.1): 75% hydrogen, 24% helium and 1% other elements.

As indicated in Sec. 1.2.1.3, the Universe experienced at redshift  $z \sim 7-10$  a period of intense reionization. This implies that the IGM at  $z \lesssim 7-10$  is a very ionized plasma. Because of the sparsity of this plasma, recombination into atoms occurs very rarely and is counterbalanced by ionization from the diffuse UV background. Taking into account only hydrogen and helium, the equations of equilibrium between charged and neutral elements of this plasma can be found in [18]:

$$\begin{aligned} \Gamma_{\text{eH}_0} n_{\text{e}} n_{\text{H}_0} + \Gamma_{\gamma\text{H}_0} n_{\text{H}_0} &= \alpha_{\text{H}^+} n_{\text{H}^+} n_{\text{e}}, \\ \Gamma_{\text{eHe}_0} n_{\text{He}_0} n_{\text{e}} + \Gamma_{\gamma\text{He}_0} n_{\text{He}_0} &= (\alpha_{\text{He}^+} + \alpha_{\text{d}}) n_{\text{He}^+} n_{\text{e}}, \\ \Gamma_{\text{eHe}^+} n_{\text{He}^+} n_{\text{e}} + \Gamma_{\gamma\text{He}^+} n_{\text{He}^+} + (\alpha_{\text{He}^{++}} + \alpha_{\text{d}}) n_{\text{He}^{++}} n_{\text{e}} &= \alpha_{\text{He}^{++}} n_{\text{He}^{++}} n_{\text{e}} + \\ &\quad \Gamma_{\text{eHe}_0} n_{\text{He}_0} n_{\text{e}} + \Gamma_{\gamma\text{He}_0} n_{\text{He}_0}, \\ \alpha_{\text{He}^{++}} n_{\text{He}^{++}} n_{\text{e}} &= \Gamma_{\text{eHe}^+} n_{\text{He}^+} n_{\text{e}} + \Gamma_{\gamma\text{He}^+} n_{\text{He}^+}, \\ n_{\text{H}} &= n_{\text{H}^+} + n_{\text{H}_0}, \\ n_{\text{e}} &= n_{\text{H}^+} + n_{\text{He}^+} + 2n_{\text{He}^{++}}, \\ Y_{\text{He}}/(4 - 4Y_{\text{He}}) &= (n_{\text{He}_0} + n_{\text{He}^+} + n_{\text{He}^{++}})/n_{\text{H}}. \end{aligned} \quad (2.2)$$

In these equations,  $\Gamma_{\text{eX}}$  and  $\alpha_{\text{X}}$ , both expressed in  $\text{cm}^3 \cdot \text{s}^{-1}$ , are respectively electronic collisional ionization and recombination rates for the X species. Both vary mainly according to the IGM temperature. The photoionization rates  $\Gamma_{\gamma\text{X}}$  for species X are expressed in  $\text{s}^{-1}$  and are calculated from a UV background model, which must be specified. Finally,  $Y_{\text{He}}$  is the helium mass abundance. These parameters and equations are notably used in cosmological simulations to model the cooling and heating rates of the IGM. The numerical solution of these equations provides the ionization state of the plasma. Their comparison with direct observation of the Ly $\alpha$  forest show that the IGM contains  $\sim 10^5$  times more ionized hydrogen on average than neutral. At the same time, measurements can be used to constrain UV background models.

The IGM temperature is a key parameter to describe its ionization state through its dependence on electronic collisional ionization and recombination rates. Over a large volume fraction of the IGM, this temperature is well described by

$$T(\vec{r}, z) = T_0(z) \left( \frac{\rho_b(\vec{r}, z)}{\bar{\rho}_b(z)} \right)^{\gamma(z)-1}, \quad (2.3)$$

where  $\rho_b$  is the baryon density. The determination of the parameters  $T_0$  and  $\gamma$  is hard and can only be done through high-resolution Ly $\alpha$  forest measurements. These measurements seem to agree that the temperature at mean density  $T_0$  increases until redshift  $z \sim 3.5$  due to the second Helium reionization. The order of magnitude of this parameter is ranging from 5,000 to 15,000 K [19].

A complete description of the properties of IGM is given in [2, 3]. IGM is also a key ingredient to understand the formation and evolution of galaxies. The star formation in galaxies depends on the amount of matter they contain: the IGM is an extensive reservoir of matter and it is responsible for the supply of matter to galaxies. From a cosmological point of view, knowing the distribution of baryons in the IGM is particularly interesting since the density of baryons traces that of dark matter. In this thesis, the observable used to trace the amount of neutral hydrogen in the IGM is the Ly $\alpha$  forest.

### 2.1.2.2 Lyman- $\alpha$ forest structure

The Lyman series is the set of hydrogen spectral lines between an excited level  $n \geq 2$  and its fundamental level. In particular, the **Lyman- $\alpha$**  (respectively **Lyman- $\beta$** ) line, noted **Ly $\alpha$**  (**Ly $\beta$** ), corresponds to the transition between the first (second) excited state and the fundamental one. In a rest frame, the wavelengths of these transitions are  $\lambda_\alpha = 1215.67 \text{ \AA}$  and  $\lambda_\beta = 1025.72 \text{ \AA}$ . When  $n$  tends to  $\infty$ , the Lyman series converges towards a limit wavelength  $\lambda_\infty = 911.75 \text{ \AA}$ , also called **Lyman limit**. Approaching  $\lambda_\infty$ , the Lyman lines are less and less separated and eventually blend together.

When a QSO is observed with a redshift  $z > 2$ , the different lines of the Lyman series can appear on its spectrum in the optical domain. Fig. 2.3 shows a QSO located at a redshift  $z = 3.42$  observed by DESI. The quasar's emission lines, Ly $\alpha$  and Ly $\beta$ , appear on this spectrum. Bluewards of the Ly $\alpha$  line and before the Ly $\beta$  line, the spectrum exhibits a series of absorption lines called Ly $\alpha$  forest. It corresponds to Ly $\alpha$  absorptions by neutral hydrogen in the IGM along the line-of-sight between the QSO and the observer.

The presence of an astrophysical signal in the Ly $\alpha$  forest was first demonstrated experimentally in 1980 [20]. Before that, James Gunn and Bruce Peterson [21] had already put forward a model to link the neutral hydrogen density of the IGM with the Ly $\alpha$  forest. We will give some tools related to this development based on the reference papers [1, 4].

### Optical depth

To describe the absorption of a physical medium, we used the framework of radiative transfer. For photons with a frequency  $\nu = c/\lambda$ , a medium is characterized by its diffusion coefficient  $\sigma(\nu)$  and its absorption opacity  $\kappa(\nu)$ . From these two properties, it is possible to define the infinitesimal **optical depth**

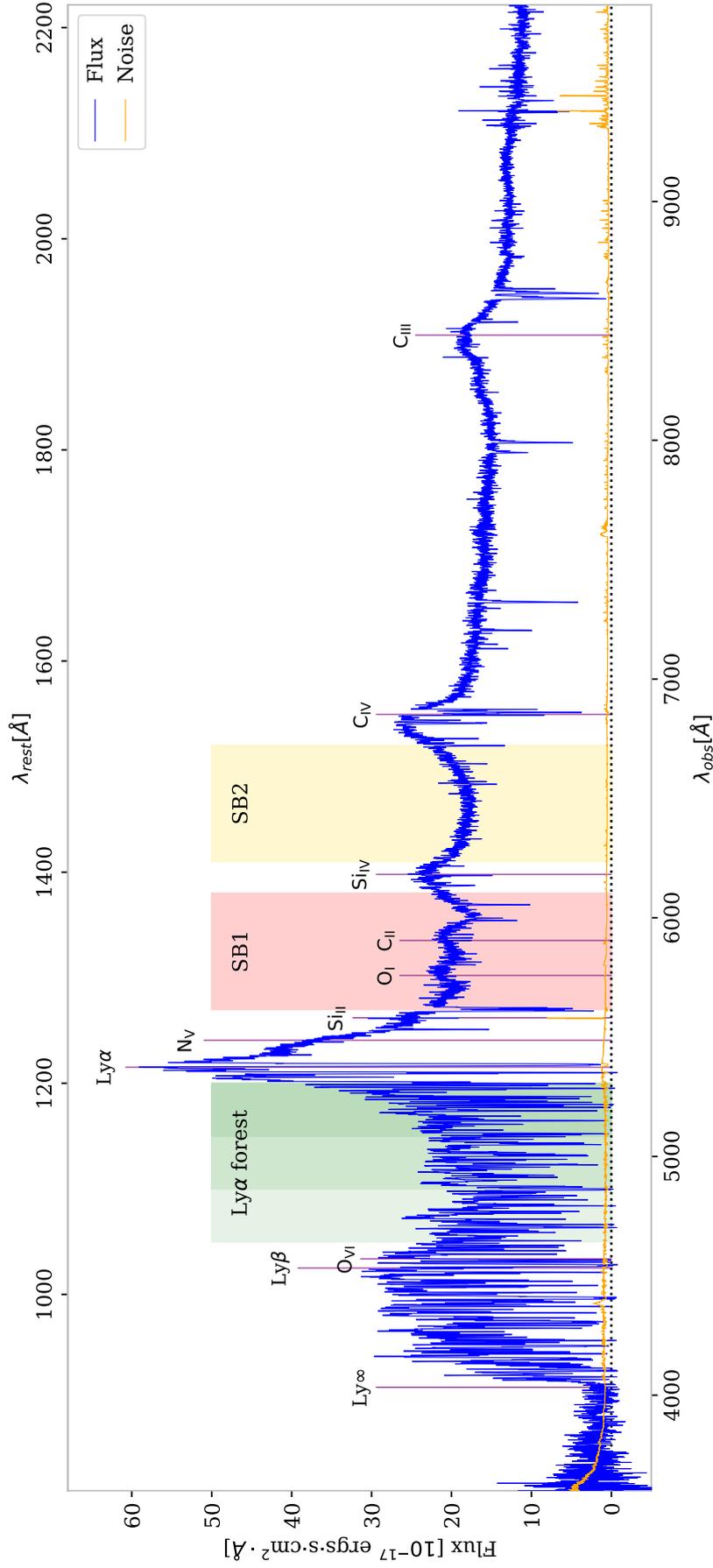


FIGURE 2.3 – The spectrum of a quasar located at a redshift  $z = 3.42$  measured by DESI with an exposure lasting 2300 seconds. This QSO was observed on the 12th April 2021 night, in the SV3 program, on the DESI tile 221 (TARGETID = 39627746095137037, RA = 217.263 $^\circ$ , DEC =  $-1.755^\circ$ ). The quasar flux is represented in blue and its instrumental noise in orange. The Ly $\alpha$  forest is shown in green. The sideband regions 1 and 2 pictured in red and yellow are used to estimate the forest contamination by metals.

$$d\tau(\nu) = (\kappa(\nu) + \sigma(\nu)) dr. \quad (2.4)$$

The optical depth  $\tau(\nu) = \int \frac{d\tau(\nu)}{dr} dr$  measures the degree of transparency of the medium. A medium can be considered as transparent to a frequency  $\nu$  when  $\tau(\nu)$  is small compared to unity. On the other hand, if this optical depth is of the same order of magnitude or greater than unity, the medium can be considered absorbent or opaque. The **fraction of transmitted flux** for a frequency  $\nu$  is given by

$$F(\nu) = \frac{f_{\text{out}}(\nu)}{f_{\text{in}}(\nu)} = e^{-\tau(\nu)}. \quad (2.5)$$

To understand the Ly $\alpha$  forest, one should thus compute the fraction of a quasar's flux absorbed by the IGM. Considering only Ly $\alpha$  absorption, it is possible to relate the opacity to the **neutral hydrogen number density**  $n_{\text{HI}}$  and the **Ly $\alpha$  scattering cross-section**  $\sigma_\alpha$ . The infinitesimal optical depth then becomes

$$d\tau_\alpha(\vec{r}, \nu) = n_{\text{HI}}(\vec{r}) \sigma_\alpha(\vec{r}, \nu) dr. \quad (2.6)$$

Since we consider a flux along a line-of-sight, it is practical to define the **column density** of neutral hydrogen

$$N_{\text{HI}} = \int n_{\text{HI}}(r_{\parallel}) dr_{\parallel}, \quad (2.7)$$

where  $r_{\parallel}$  is the projection of  $\vec{r}$  along the quasar line-of-sight. The column density is then associated to a quasar sky position.

### Modeling the Ly $\alpha$ cross-section

Around the Ly $\alpha$  resonance, the interaction cross-section between neutral hydrogen atoms at rest and photons with frequency  $\nu$  is given by a Lorentz line profile:

$$\sigma_{\text{rest},\alpha}(\nu) = \left( \frac{\pi e^2}{m_e c} \right) \left[ \frac{1}{4\pi\epsilon_0} \right] f_\alpha \frac{\Gamma_\alpha/4\pi^2}{(\nu - \nu_\alpha)^2 + (\Gamma_\alpha/4\pi)^2}, \quad (2.8)$$

where  $f_\alpha$  is the oscillator strength of the Ly $\alpha$  transition and the radiation damping term  $\Gamma_\alpha$  is given by

$$\Gamma_\alpha = \frac{g_1 f_\alpha}{g_2} \left[ \frac{1}{4\pi\epsilon_0} \right] \frac{8\pi^2 e^2}{m_e c \lambda_\alpha^2}. \quad (2.9)$$

Here,  $g_1$  and  $g_2$  correspond to ground state and first excited state statistical weights. This Lorentz profile takes into account the energy spread due to the lifetime broadening of the excited state which is linked to the Heisenberg uncertainty principle.

The atoms of the IGM are not at rest. To take into account the temperature of the IGM, it is necessary to add **thermal broadening** to the line profile. Ignoring non-thermal velocities in the hydrogen gas, thermal broadening is quantified by the Doppler parameter  $b_{\text{th}}$  which corresponds to thermal velocity:

$$b_{\text{th}} = \sqrt{\frac{2k_{\text{B}}T}{m_{\text{HI}}}}. \quad (2.10)$$

The line profile, which includes the thermal motion, is obtained by convolving Eqn. 2.8 by a Maxwellian function. The final line profile is then given by a Voigt-Hjerting function, defining  $a = \Gamma_{\alpha}/(4\pi\Delta\nu_{\text{D}})$  the ratio between the damping width and  $\Delta\nu_{\text{D}} = b_{\text{th}}/\lambda_{\alpha}$  which is the broadening width, and  $x = (\nu - \nu_{\alpha})/\Delta\nu_{\text{D}}$  the frequency shift from line center, this function reads:

$$V(a, x) = \frac{a}{\pi} \int_{-\infty}^{\infty} dy \frac{e^{-y^2}}{(x-y)^2 + a^2} = \begin{cases} \sim e^{-x^2} [\exp(a^2) \operatorname{erfc}(a)] \sim e^{-x^2}, & \text{for core region,} \\ \sim \frac{a}{\sqrt{\pi}x^2}, & \text{for damping wings.} \end{cases} \quad (2.11)$$

The thermally averaged Ly $\alpha$  cross-section is given by

$$\sigma_{\alpha}(\nu) = \left( \frac{\pi e^2}{m_e c} \right) \left[ \frac{1}{4\pi\epsilon_0} \right] f_{\alpha} \frac{V(a, x)}{\pi^{1/2} \Delta\nu_{\text{D}}}. \quad (2.12)$$

As indicated in Eqn. 2.11, this profile can be decomposed as a Gaussian core with damping wings. For the less dense absorption systems of the IGM, damping wings can be neglected, and the line profile is Gaussian.

### Gunn-Peterson effect

To obtain Eqn. 2.12, only the passage through a narrow portion of the IGM was considered. To fully describe the IGM absorptions, one must consider all absorbers along the QSO line-of-sight, taking into account the expansion of the Universe.

As illustrated in Fig. 2.4, the Ly $\alpha$  forest results from the accumulation of all IGM absorbers along the line-of-sight. Due to the expansion of the Universe, these absorbers will be associated to different redshifts. For an absorber located at a redshift  $z_{\text{abs}}$ , the QSO spectrum will show an absorption line located at wavelength

$$\lambda = (1 + z_{\text{abs}}) \lambda_{\alpha}. \quad (2.13)$$

The addition of all the absorber contributions to the Ly $\alpha$  forest is called the Gunn-Peterson effect [21]. Similarly, forests for other lines are created by the same effect. Thus, we can define in the same way the Ly $\beta$  forest and the forest related to the other elements (metals) present in the IGM. For example, in Fig. 2.3, the absorption features located bluewards of the Ly $\beta$  line results from the combination of the Ly $\beta$  and Ly $\alpha$  forests.

To compute the Ly $\alpha$  forest as a function of frequency  $\nu$ , we must integrate the infinitesimal optical depth defined by in Eqn. 2.6 along the line-of-sight, for all absorbers positioned in  $r$  and verifying Eqn. 2.13. We thus obtain:

$$\tau_{\alpha}(\nu) = \int n_{\text{HI}}(r) \sigma_{\alpha}(r, \nu) \delta_{\text{D}} \left( 1 + z_{\text{abs}} - \frac{\nu_{\alpha}}{\nu} \right) dr. \quad (2.14)$$

For absorbers with the characteristic column density of the Ly $\alpha$  forest ( $N_{\text{HI}}$  between  $10^{12}$  and  $10^{16} \text{ cm}^{-2}$ ), the absorption profile can be simplified to a Gaussian function. Objects with higher column density will be treated in Sec. 2.1.3.

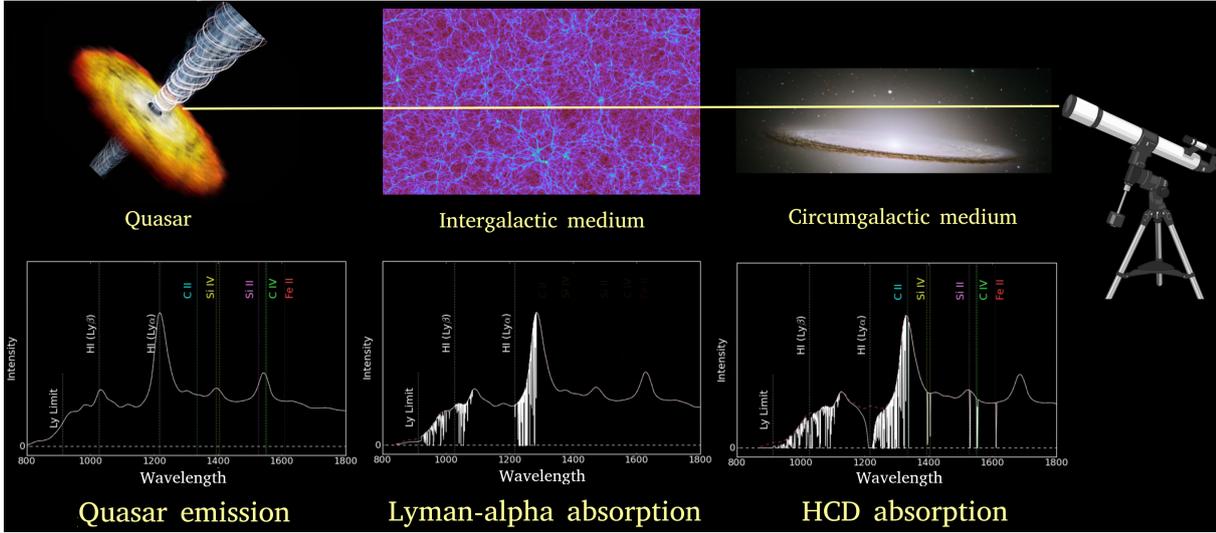


FIGURE 2.4 – Schematic representation of the formation of the Ly $\alpha$  forest adapted from A. Pontzen video. The light emitted by the quasar is absorbed by the neutral hydrogen of the IGM, forming the Ly $\alpha$  forest. A high concentration of neutral hydrogen, called a high column density region, creates a saturated absorption. This can be the case when crossing the neighborhood of a galaxy.

The expansion of the Universe is taken into account using comoving coordinates. The physical distance  $r$  is transformed to Hubble flow velocity by the coordinate change  $dr = dv/H(v)$ . As explained in Sec. 1.1.3, the redshift of the absorber  $z_{\text{abs}}$  is the combination of the Hubble flow and its particular velocity projected along the line-of-sight  $v_{\parallel}$ . The peculiar velocities create a blending of the Ly $\alpha$  absorption lines and produce redshift space distortions (RSD).

The IGM temperature was assumed constant until now, whereas it depends on the position. We correct this by adding a dependence on the redshift in the Doppler parameter  $b(z)$ .

We place ourselves in the Hubble flow velocity space. It is possible to link this velocity  $v$  to the frequency  $\nu$  by the relation  $v = c(\nu_{\alpha}/\nu - 1)$ . Finally, the optical depth can be expressed as:

$$\tau_{\alpha}(v) = \left( \frac{\sqrt{\pi} e^2}{m_e} \right) \left[ \frac{1}{4\pi\epsilon_0} \right] \lambda_{\alpha} f_{\alpha} \int n_{\text{HI}}(v') \frac{1}{b(v') H(v')} e^{-\left( \frac{v-v'-v_{\parallel}(v')}{b(v')} \right)^2} dv'. \quad (2.15)$$

This equation gives the fraction of transmitted flux along a line-of-sight. A calculation of this optical depth requires the knowledge of speed, temperature and density of neutral hydrogen at all positions along the line-of-sight. For example, this equation is used to compute Ly $\alpha$  observables from hydrodynamical simulations.

### Fluctuating Gunn-Peterson approximation

The fluctuating Gunn-Peterson approximation (FGPA) [22, 23, 24, 25, 26] is a useful approximation to quickly compute the Ly $\alpha$  forest on large scales. First, let us consider the Gunn-Peterson effect for a uniform IGM. In this non-physical case, the absorption lines blend, and the optical depth becomes

$$\tau_{\alpha}^{\text{u}} = \frac{\pi e^2}{m_e c} \frac{f_{\alpha} \lambda_{\alpha}}{H(z)} n_{\text{HI}}. \quad (2.16)$$

The first approximation is to consider only hydrogen in the ionization process expressed by Eqn. 2.2. Assuming additionally that the IGM is in equilibrium and highly photoionized, we can express  $n_{\text{HI}}$  as a function of the photoionization rate  $\Gamma$  and the recombination rate  $\alpha(T)$  of hydrogen:

$$n_{\text{HI}} = \frac{n_{\text{H}}n_{\text{e}}\alpha(T)}{\Gamma} = \frac{1.16n_{\text{H}}^2\alpha(T)}{\Gamma}. \quad (2.17)$$

We place ourselves in the framework of the optically thin approximation  $\tau_{\alpha}^{\text{u}} \ll 1$ . For typical IGM temperatures ( $T_0 \sim 10^4$  K), this approximation implies that the recombination rate is a function of temperature  $\alpha \propto T^{-0.7}$ . Assuming primordial hydrogen and helium abundances, it is possible to relate the number density of hydrogen  $n_{\text{H}}$  to the baryon density parameter  $\Omega_{\text{b}}$ . A full calculation [22] yields the following expression for the uniform optical depth:

$$\tau_{\alpha}^{\text{u}} = 2.31 \times 10^{-4} (1+z)^6 \left( \frac{\Omega_{\text{b}} h^2}{0.0125} \right)^2 \left( \frac{T}{10^4 \text{ K}} \right)^{-0.7} \left( \frac{\Gamma}{10^{-12} \text{ s}^{-1}} \right)^{-1} \left( \frac{H(z)}{100 \text{ km} \cdot \text{s}^{-1} \cdot \text{Mpc}^{-1}} \right)^{-1}. \quad (2.18)$$

In a second phase, we consider an extension to the uniform model by adding slowly-varying small perturbations to the IGM density and temperature. The thermal broadening and peculiar velocities are ignored. The latter approximation is valid when considering regions of the IGM that are not very dense, and when considering large scales (larger than  $\sim 2 h^{-1} \cdot \text{Mpc}$  [24]). Using these approximations, we can link the optical depth to the one obtained in the uniform IGM case:

$$\tau_{\alpha} = \tau_{\alpha}^{\text{u}} \left( \frac{n_{\text{H}}}{\bar{n}_{\text{H}}} \right)^2 \frac{\alpha(T)}{\alpha(T_u)}, \quad (2.19)$$

where  $T_u$  is the temperature of the uniformly distributed IGM. Finally, we assume that the temperature dependence on density follows Eqn. 2.3 with fixed parameters  $\gamma$  and  $T_0$ . We obtain:

$$\tau_{\alpha} \propto \left( \frac{1+z}{4} \right)^6 \left( \frac{\Omega_{\text{b}} h^2}{0.0125} \right)^2 \left( \frac{T_0}{10^4 \text{ K}} \right)^{-0.7} \left( \frac{\Gamma}{10^{-12} \text{ s}^{-1}} \right)^{-1} \left( \frac{H(z)}{100 \text{ km s}^{-1} \text{Mpc}^{-1}} \right)^{-1} (1 + \delta_{\text{b}})^{\beta}, \quad (2.20)$$

where  $\delta_{\text{b}}$  is the baryon density contrast and  $\beta = 2 - 0.7(\gamma - 1)$ .

Within the domain of validity of this approximation, the Ly $\alpha$  forest provides a power-law relation between the Ly $\alpha$  optical depth and the underlying baryon fluctuations. This is why the Ly $\alpha$  forest is a powerful tool to probe large-scale cosmological structures.

### 2.1.3 Astrophysical systematics of Lyman- $\alpha$ forest

However, the Ly $\alpha$  forest signal is contaminated by several systematics that come from additional astrophysical effects. In general, systematics are undesirable effects which generate uncertainty errors in measurements. One can classify them in three categories: **observational systematics** related to contamination by the observational method, **analysis systematics** related to the quality of the methods used for the data analysis, and **astrophysical systematics** which are intrinsic to the observable used. This section will explain the astrophysical systematics of the Ly $\alpha$  forest. Their identification and correction are essential steps to interpret the data.

### High column density objects

The first significant contaminants of the Ly $\alpha$  forest are the **high column density (HCD)** objects. As the QSO light travels across the IGM, it crosses regions of high neutral hydrogen concentration. Those are interpreted as gas clouds dense enough to be shielded from ionizing radiation but not dense enough for star formation. This is the case for regions in the periphery of galaxies which are called circumgalactic medium (CGM).

As shown in Fig. 2.4, the passage of the QSO light in the CGM leads to a saturated ( $F = 0$ ) and broad absorption. In this framework, the approximations made to compute the Ly $\alpha$  forest optical depth in Eqn. 2.15 are no longer valid. For HCD objects, the lifetime broadening dominates the thermal one. It leads to large damping wings [27], independent of the absorber's velocity profile. Consequently, crossing a small dense region of the Universe will impact a large portion of the spectrum. It implies suppressing the link between the Ly $\alpha$  forest and the matter density on a wide range of wavelengths.

An HCD object is categorized by its column density defined by Eqn. 2.7 and can be placed in different categories. Here, the names follow the conventions used in [27, 28]:

- **Damped Lyman- $\alpha$  systems (DLA)**: They are the densest absorber, with a column density  $N_{\text{HI}} > 10^{20.3} \text{ cm}^{-2}$ . They contaminate the Ly $\alpha$  forest signal. They are studied independently to characterize the CGM and constrain galaxy formation models.
- **Super Lyman limit systems (sLLS)**: Intermediate features with a column density defined by  $10^{19} < N_{\text{HI}} < 10^{20.3} \text{ cm}^{-2}$ . They are attributed to the densest parts of the galaxy halos.
- **Lyman limit systems (LLS)**: The least dense HCDs with a column density  $10^{17.2} < N_{\text{HI}} < 10^{19} \text{ cm}^{-2}$ . These systems are difficult to disentangle from Ly $\alpha$  forest absorptions.

The Ly $\alpha$  forest, properly speaking, corresponds to all features that have a column density  $N_{\text{HI}} < 10^{16} \text{ cm}^{-2}$ . This column density is insufficient to completely saturate the absorption at redshifts where the Ly $\alpha$  forest is mostly used ( $2 < z < 5$ ).

Taking these HCDs into account is an essential part of any Ly $\alpha$  study. They can be corrected with theoretical forward models, which is difficult because the CGM and neighboring galaxy profiles must be modeled. Another strategy used in this thesis is to mask these regions and correct the profile of the damping wings.

### Metal absorptions

As seen previously, there is an associated forest for each absorbing element present in the IGM. The Ly $\alpha$  forest has the most prominent signal, but metals also contaminate it when their transition is at a higher wavelength than the Ly $\alpha$  emission line.

As shown in detail in Fig. 2.3, the transitions contaminating the Ly $\alpha$  forest are NV, SiII, OI, CII, SiIV, CIV, CIII and other minor metals. These metal forests can be used separately as an additional tracer of the matter. For Ly $\alpha$  studies, it is necessary to consider the share of absorptions due to metals within the Ly $\alpha$  forest. The elements SiII, SiIV and CIV impact the most the Ly $\alpha$  forest. Their impact can be alleviated by modeling or statistical subtraction using other spectral regions.

### Broad absorption line QSOs

The last astrophysical contaminants are the **broad absorption line (BAL)** quasars. The BAL quasars are a subclass of QSOs with broad absorptions blueshifted from QSO emissions

by at least  $2000 \text{ km}\cdot\text{s}^{-1}$  and with velocity widths of at least  $2000 \text{ km}\cdot\text{s}^{-1}$  [29]. These QSOs represent nearly 10% of the total population.

In general, an absorption is observed near the CIV emission line. However, other absorptions can be next to all major metal lines in a QSO spectrum. The characterization of these objects is done by calculating their balnicity index [30, 31] which is given by

$$BI = - \int_{25,000}^{3,000} [1 - f(v)/0.9] C dv, \quad (2.21)$$

where  $v$  is the shift in speed to the considered emission line,  $f(v)$  the flux at this position and  $C$  is a Boolean value which is zero if the two BAL conditions are not satisfied.

The most accepted explanation is that BALs originate from absorptions by the QSO torus or by regions close to this torus (See Fig. 2.1). As these absorptions reflect the immediate environment of the quasar, they contaminate the Ly $\alpha$  forest, which focuses on the IGM absorptions. Moreover, the absorptions of the BAL QSOs being rather broad and deep can strongly bias the signal of the Ly $\alpha$  forest. In general, the spectra of BAL QSOs are discarded for Ly $\alpha$  analysis.

## 2.2 Clustering with the Lyman- $\alpha$ forest

In the previous section, we highlighted that the Ly $\alpha$  forest traces the amount of neutral hydrogen along the line-of-sight of a QSO. The observation of the Ly $\alpha$  forest can be used in astrophysics to characterize the properties of the IGM. In the context of this thesis, the Ly $\alpha$  forest is employed in view of cosmological applications. Here, we present basic statistical observables in this context.

### 2.2.1 Lyman- $\alpha$ forest correlations

The calculation of correlations with the Ly $\alpha$  forest fits into the general framework defined in Sec. 1.3.1. However, some important nuances to Ly $\alpha$  analysis must be mentioned.

#### Ly $\alpha$ contrast

Our starting point observable is the flux  $f$  represented in Fig. 2.3. Our goal is to measure the fraction of transmitted flux contrast of the Ly $\alpha$  forest. From now, we will call this observable as the **Ly $\alpha$  contrast**. It is given by

$$\delta_F(\lambda) = \frac{F(\lambda)}{\bar{F}(\lambda)} - 1, \quad (2.22)$$

where  $F$  is the fraction of transmitted flux given by Eqn. 2.5, and  $\bar{F}(\lambda) = \bar{F}(z)$  the average value of this variable at redshift  $z$ .

Measuring this contrast from the observed spectra requires an additional step. Indeed, it is necessary to estimate the quasar continuum to recover  $F$ . Thus, the fraction of transmitted flux is replaced in Eqn. 2.22 by the ratio of the observed flux to the quasar continuum:

$$\delta_F(\lambda) = \frac{f(\lambda)}{C_q(\lambda, z_q)\bar{F}(\lambda)} - 1. \quad (2.23)$$

The continuum is a different function for each QSO  $q$ . This function depends on the observed wavelengths  $\lambda$  and the QSO rest-frame wavelengths  $\lambda_{\text{rf}} = \lambda/(1+z_q)$  through the QSO redshift  $z_q$ . The first step of a Ly $\alpha$  study is to estimate the QSO continuum from the flux. This step, called **continuum fitting** or **Ly $\alpha$  contrast extraction**, can be done in several ways. The approach used in this thesis will be presented in Sec. 5.1.2. The representation of a QSO continuum and of the product  $C_q(\lambda, z_q)\bar{F}(\lambda)$  is given in Fig. 2.5 for the same spectrum as in Fig. 2.3.

#### Correlations

From the measurement of the Ly $\alpha$  contrast, it is possible to define the two-point correlation function (or auto-correlation)  $\xi_\alpha$  and the two-point cross-correlation function (also expressed cross-correlation) with another tracer  $Y$ , noted  $\xi_{\alpha Y}$ . These two functions are given by

$$\xi_\alpha(\vec{r}) = \langle \delta_F(\vec{x})\delta_F(\vec{x} + \vec{r}) \rangle_{\vec{x}}, \quad (2.24)$$

$$\xi_{\alpha Y}(\vec{r}) = \langle \delta_F(\vec{x})\delta_Y(\vec{x} + \vec{r}) \rangle_{\vec{x}}. \quad (2.25)$$

The  $Y$  tracer used can be quasars or voids, for example. A measurement of the cross-correlation between the Ly $\alpha$  forest and cosmic voids is the main subject of Chap. 7.

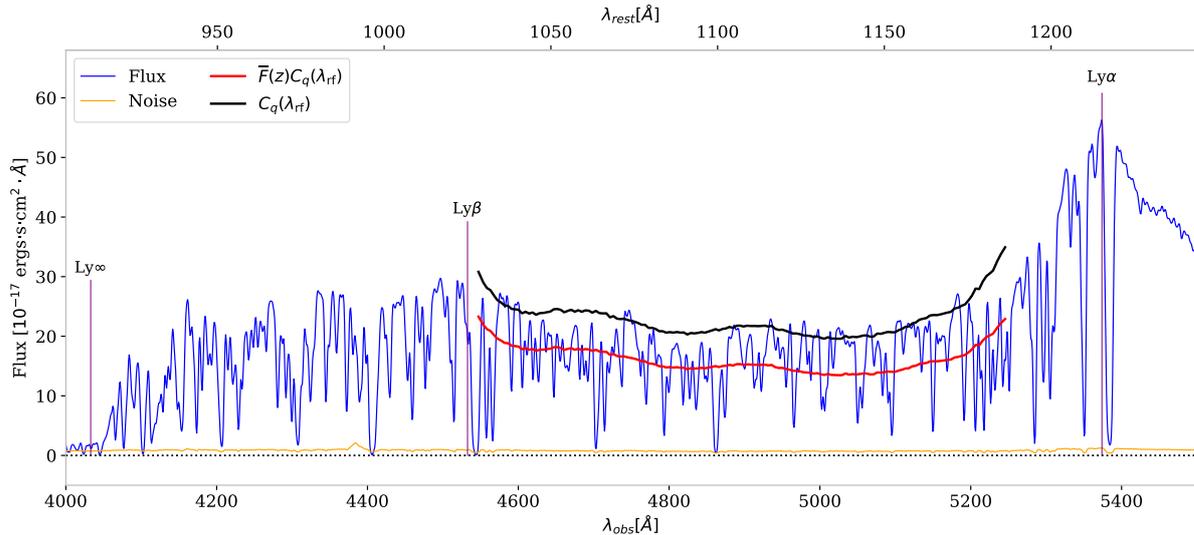


FIGURE 2.5 – Zoom in on the quasar spectrum shown in Fig. 2.3. Here, the flux is Gaussian smoothed with a  $1.2 \text{ \AA}$  kernel. The quasar continuum is estimated by a polynomial fit of the  $C_q(\lambda, z_q)\bar{F}(\lambda)$  term represented in red. This method is used in my thesis and will be detailed in Sec. 5.1.2. The continuum is computed from an average fraction of transmitted flux  $\bar{F}(\lambda)$  computed using [32].

The Ly $\alpha$  forest can also be modeled in Fourier space. The **three-dimensional Ly $\alpha$  forest power spectrum**  $P_\alpha$ , which is the Fourier transform of the auto-correlation as indicated by Eqn. 1.70, is then defined by

$$P_\alpha(k)\delta_D^{(3)}(\vec{k} + \vec{k}') = \frac{1}{(2\pi)^3} \langle \delta_F(\vec{k})\delta_F(\vec{k}') \rangle. \quad (2.26)$$

The Ly $\alpha$  forest is a cosmological tracer with a particular geometry, as it is strongly anisotropic. Therefore, it is interesting to focus on the study of correlations along the line-of-sight. This is the case when measuring the **one-dimensional Ly $\alpha$  forest power spectrum**  $P_{1D,\alpha}$  which can be linked to  $P_\alpha$  such that:

$$P_{1D,\alpha}(k_{\parallel}) = \int \frac{d\vec{k}_{\perp}}{(2\pi)^2} P_\alpha(\vec{k}). \quad (2.27)$$

The three-dimensional power spectrum is the observable measured on the simulations in Chap. 3 and the one-dimensional power spectrum is the main measurement of Chap. 5.

## 2.2.2 Modeling of bias and redshift space distortions

To simply describe the behavior of the Ly $\alpha$  contrast, we can use a linear model which relates this contrast to the matter density. For that, we decompose the Ly $\alpha$  contrast in a long and a short wavelength components  $\delta_F = \delta_{F,L} + \delta_{F,S}$  following [33]. We focus here on the large-scale modes variations at least superior to about  $2 h^{-1} \cdot \text{Mpc}$ , such that for this section  $\delta_F = \delta_{F,L}$ .

### Ly $\alpha$ bias

As shown in previous section with Eqn. 2.5, the Ly $\alpha$  contrast is the exponential of the optical depth which is related to the baryon density (Eqn. 2.20). Consequently, the Ly $\alpha$  forest is a non-linear tracer of baryons, which is itself a tracer of matter using the results of Sec. 1.3. In a general way, we can relate the contrast of an observable that traces the matter to the matter contrast  $\delta_m$  by some function:

$$\delta_F = f(\delta_m). \quad (2.28)$$

A linear relation between both contrasts is expected for the large-scale, linear modes. Thus, it is possible to define the linear bias called **optical depth bias**  $b_{\tau_\alpha}$  and **Ly $\alpha$  bias**  $b_\alpha$ :

$$\begin{aligned} b_{\tau_\alpha} &= \left. \frac{\partial \delta_{\tau_\alpha}}{\partial \delta_m} \right|_{\delta_{\tau_\alpha}=0}, \\ b_\alpha &= \left. \frac{\partial \delta_F}{\partial \delta_m} \right|_{\delta_F=0}. \end{aligned} \quad (2.29)$$

Again considering the large-scale modes and assuming linearity, we obtain the relation  $\delta_F = b_\alpha \delta_m$ . It is important to note that at the heart of an overdensity ( $\delta_m \gg 1$ ), the Ly $\alpha$  absorption is maximal ( $F = 0$ ), and consequently the Ly $\alpha$  contrast is equal to  $-1$ . On the contrary, in the absence of matter ( $\delta_m = -1$ ), the Ly $\alpha$  contrast is equal to  $(1/\bar{F}) - 1 > 0$ . Therefore, the Ly $\alpha$  bias must be negative to link these two contrasts.

The Ly $\alpha$  forest traces the IGM regions with a typical density similar to the average density of the Universe. The bias between neutral hydrogen and the total matter is very close to unity for these regions. However, the Ly $\alpha$  forest is non-linear by construction as it is expressed in Eqn. 2.5. This relation "compresses" the variations of  $\tau_\alpha$  and makes the absolute value of the Ly $\alpha$  bias smaller than unity. It is typically between 0 and 1 which is much lower than the bias of the Boolean tracers (LRG, ELG, or QSOs as tracers). As an example, one of the last Ly $\alpha$  bias measurement is  $b_\alpha = -0.2014 \pm 0.0032$ , at  $z = 2.33$  [34].

### Redshift space distortions

In practice, the Ly $\alpha$  contrast is observed in the redshift space introduced in Sec. 1.3.3.1. We will now adapt the general RSD relation given by Eqn. 1.140 to the large-scale modes of the Ly $\alpha$  forest. Another derivation of RSD will be given in the particular case of voids in Chap. 7.

If we consider a point located at physical position  $\vec{r}$ , its position in redshift space  $\vec{s}$  is given by modifying the RSD relation of Eqn. 1.135 to the case of a fixed observer without wide-angle effects:

$$\vec{s} = \vec{r} + (1+z) \frac{\hat{r} \cdot \vec{u}}{H(z)} \hat{r}, \quad (2.30)$$

where  $\vec{u}$  is the physical velocity and the normalized vector  $\hat{r} = \vec{r}/r$  gives the radial direction.

For Boolean tracers, Eqn. 1.140 is derived by making use of the tracer number conservation between physical and redshift space. In the case of the Ly $\alpha$  forest, it is the optical depth that is conserved by the transformation 2.30. We thus obtain a relation between the optical depth expressed in the redshift space  $\tau_\alpha^s$  and that expressed in the physical space  $\tau_\alpha$ :

$$\tau_\alpha(\vec{r}) d^3 \vec{r} = \tau_\alpha^s(\vec{s}) d^3 \vec{s} = \tau_\alpha^s(\vec{r}) \left| \frac{d^3 \vec{s}}{d^3 \vec{r}} \right| d^3 \vec{r}. \quad (2.31)$$

The Jacobian of this transformation is given by

$$\left| \frac{d^3 \vec{s}}{d^3 \vec{r}} \right| = 1 + \frac{1}{\mathcal{H}} \frac{\partial \vec{u} \cdot \hat{r}}{\partial r_{\parallel}}, \quad (2.32)$$

where the radial component of the position is defined by  $r_{\parallel} = \vec{r} \cdot \hat{r}$ .

As we focus on the large-scale modes, we can use the relation expressing the velocity field under the assumption of linearity given by Eqn. 1.130. After integrating into Fourier space, this relation is used to calculate the value of the Jacobian in Eqn. 2.32. Thus, by replacing it in Eqn. 2.31, we derive the following relation in Fourier space:

$$\tau_{\alpha}^s(\vec{k}) = \tau_{\alpha}(\vec{k})(1 + f\mu^2\delta_m(\vec{k})). \quad (2.33)$$

According to Eqn. 2.5, the fraction of transmitted flux is a functional such that  $F = \exp(-\tau_{\alpha}^s)$ . Following [33], we carry out a Taylor expansion of this functional and cut terms higher than linear order since we consider the large-scale modes:

$$\begin{aligned} F[\tau_{\alpha}^s(\vec{k})] &= \sum_{n=0}^{\infty} \frac{F^{(n)}[0] \left( \tau_{\alpha}^s(\vec{k}) - 0 \right)^n}{n!} \\ &= \sum_{n=0}^{\infty} \frac{F^{(n)}[0] \tau_{\alpha}(\vec{k})^n}{n!} \left( 1 + f\mu^2\delta_m(\vec{k}) \right)^n \\ &= F^{(0)}[0] + F^{(1)}[0] \tau_{\alpha}(\vec{k}) \left( 1 + f\mu^2\delta_m(\vec{k}) \right), \end{aligned} \quad (2.34)$$

where  $F^{(n)} = d^n F / d\tau_{\alpha}^n$  is the  $n^{\text{th}}$  derivative of  $F$  with respect to  $\tau_{\alpha}$ .

We apply the same Taylor decomposition for the fraction of transmitted flux in physical space:

$$F[\tau_{\alpha}(\vec{k})] = F^{(0)}[0] + F^{(1)}[0] \tau_{\alpha}(\vec{k}). \quad (2.35)$$

We define the **velocity bias**  $b_{\eta}$  by the following expression:

$$b_{\eta} = \tau_{\alpha} \left. \frac{\partial \delta_F}{\partial \tau_{\alpha}} \right|_{\tau=0} = \tau_{\alpha} \frac{F^{(1)}[0]}{\bar{F}}. \quad (2.36)$$

By replacing Eqn. 2.36 and 2.35 in Eqn. 2.34, the redshift-space fraction of transmitted flux can be expressed as

$$F[\tau_{\alpha}^s(\vec{k})] = F[\tau_{\alpha}(\vec{k})] + b_{\eta} \bar{F} f \mu^2 \delta_m(\vec{k}). \quad (2.37)$$

In this expression,  $b_{\eta}$  links  $F[\tau_{\alpha}^s(\vec{k})]$  to  $f\mu^2\delta_m(\vec{k}) \propto \partial \vec{u} \cdot \hat{r} / \partial r_{\parallel}$ . The latter term corresponds to the velocity gradient, which clarifies the name given to  $b_{\eta}$ . Finally, we obtain a relation between the Ly $\alpha$  and matter density contrasts by using the definitions in Eqn. 2.22 and 2.29:

$$\delta_F^s(\vec{k}) = b_{\alpha} \left( 1 + \beta_{\alpha} \mu^2 \right) \delta_m(\vec{k}). \quad (2.38)$$

In this relationship, the  $\beta_{\alpha}$  parameter also called **RSD parameter** is defined as

$$\beta_{\alpha} = \frac{b_{\eta} f}{b_{\alpha}}. \quad (2.39)$$

Unlike Boolean tracers, this parameter contains an additional bias  $b_\eta$ . Consequently, the knowledge of  $b_\eta$  is needed to characterize the logarithmic growth rate of linear perturbations  $f$  from a  $\beta_\alpha$  measurement. Finally, using the definition of the three-dimensional Ly $\alpha$  power spectrum given by Eqn. 2.26, we obtain:

$$P_\alpha(\vec{k}, z) = b_\alpha^2(z) \left(1 + \beta_\alpha(z)\mu^2\right)^2 P_m(k, z), \quad (2.40)$$

where  $P_m$  is the linear matter power spectrum defined in Sec. 1.3.2.2. We will use this relation to calculate the Ly $\alpha$  bias in the simulations of Chap. 3.

## 2.3 Some applications of the Lyman- $\alpha$ forest in cosmology

In this section, I present some of the main Ly $\alpha$  forest observations, and related cosmological applications, in the global context introduced in Sec. 1.4. I focus here on measurements most connected to my thesis work.

### 2.3.1 Observation types

For Ly $\alpha$  studies, one may conveniently distinguish the type of observation performed according to the resolution  $R = \Delta\lambda/\lambda$  of the spectrograph used (see Sec. 4.1 for more details). Some observations called **high-resolution** focus on studying a small number of QSOs over a long exposure time with both a very high resolution ( $30,000 < R < 100,000$ ) and high signal-to-noise ratio (mean SNR  $\sim 20$ ). Conversely, the so-called **moderate-resolution** observations study a large amount of spectra with a lower resolution ( $2,000 < R < 10,000$ ) and signal-to-noise ratio (mean SNR  $\sim 2$ ).

#### High-resolution observations

This type of observations was the first to provide spectra usable for both astrophysics and cosmology.

The European Southern Observatory (ESO)'s Very Large Telescope (VLT) located in Chile is equipped with several spectrographs allowing the observation of the Ly $\alpha$  forest. The SQUAD survey, whose last release is given in [36], has collected 467 spectra using the Ultraviolet and Visual Echelle Spectrograph (UVES). This spectrograph has a resolution of  $40,000 < R < 110,000$ . A survey with a more moderate resolution ( $4,000 < R < 17,000$ ) called XQ-100 [37] used the XSHOOTER spectrograph installed on the VLT.

The KODIAQ [35, 38] survey measured 831 Ly $\alpha$  forest with a resolution  $36,000 < R < 103,000$  using the High-Resolution Echelle Spectrograph (HIRES). This instrument is installed on the Keck telescope in Hawaii. An example of KODIAQ high-resolution spectra is given in Fig. 2.6.

Finally, the Cosmic Origins Spectrograph (COS) [39] was installed on the Hubble Space Telescope (HST). This spectrograph measured 82 AGN spectra with the resolution  $12,000 < R < 24,000$ . Space telescopes can see the Ly $\alpha$  forest in the UV range ( $z < 2.0$ ), unlike the ground-based observations for which this wavelength range is absorbed by the atmosphere.

The ArmazoNes high Dispersion Echelle Spectrograph (ANDES) is currently in construction on the ESO's Extremely Large Telescope (ELT). Due to the colossal width of the ELT's primary mirror (39 m), the integration time will be significantly reduced for objects with fixed SNR. Consequently, ANDES is expected to provide a large amount of high-resolution spectra ( $R < 100,000$ ).

#### Moderate-resolution observations

The observation of a large number of QSO spectra with a medium resolution was made possible with the use of multi-object spectrographs, which are the main subject of Chap. 4. Tab. 2.1 provides the main characteristics of the surveys detailed here.

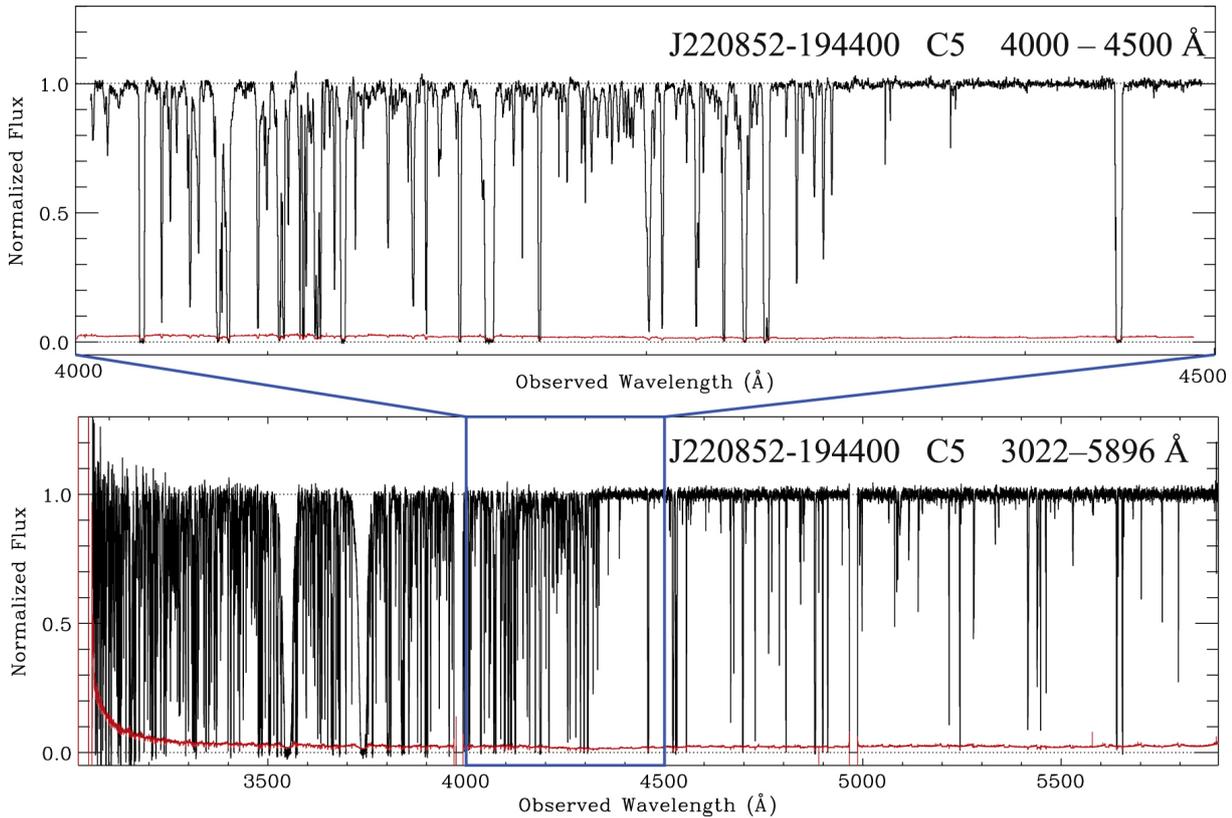


FIGURE 2.6 – Example of high-resolution spectra normalized by its continuum taken from [35]. This QSO, called J220852-194400, is located at redshift  $z = 2.573$ . The noise is shown in red, and the top panel is a zoomed-in version of the bottom one.

Survey	Date	Resolution	Density [ $\text{deg}^{-2}$ ]	Area [ $\text{deg}^2$ ]
SDSS/BOSS/eBOSS	1998 - 2019	1,500 - 2,700	21	10,000
DESI	2020 -	2,000 - 5,100	60	14,000
WEAVE-QSO MR1	2022 -	4,000 - 7,500	48	6,000
WEAVE-QSO MR2	2022 -	4,000 - 7,500	111	418
WEAVE-QSO HR	2022 -	16,000 - 23,000	4	6,000
4MOST MR	2024 -	4,000 - 7,800	-	17,000
4MOST HR	2024 -	19,000 - 21,000	-	17,000

TABLE 2.1 – Operation date, spectral resolution, measured or expected Ly $\alpha$  forest QSO density and area covered by the eBOSS [40], DESI [41, 42], WEAVE-QSO [43] and 4MOST [44] surveys. WEAVE-QSO numbers are from a conference presentation (Moriond 2022). HR and MR abbreviations respectively mean high and moderate resolutions. The observation plans of 4MOST are not yet fully determined.

The first survey to observe a large number of the Ly $\alpha$  forest was the Sloan Digital Sky Survey (SDSS) [45] with more than 3,000 Ly $\alpha$  QSOs. Surveys using the SDSS telescope like BOSS [46] and subsequently its extension eBOSS [40] increased drastically this amount to 200,000 Ly $\alpha$  forest. The properties of these two surveys will be detailed in Sec. 4.2.

Two surveys DESI [41, 42] and WEAVE-QSO [43] will also perform massive observations of Ly $\alpha$  forest. DESI is currently taking data with a moderate-resolution spectrograph and will observe almost a million Ly $\alpha$  forest QSOs. This survey, as well as its Ly $\alpha$  forest program, are presented in Sec. 4.3. The WEAVE-QSO survey is composed of moderate- and high-resolution spec-

trographs, and its first light is expected in mid-2022. WEAVE is being installed on the William Herschel Telescope in La Palma and will simultaneously observe 1000 moderate-resolution (MR in Tab. 2.1) and 1000 high-resolution (HR) spectra with dedicated fibers. As detailed in Tab. 2.1, a high quasar density program (MR2), as well as DESI-like observations (MR1), are planned. These two surveys will significantly improve the contribution of Ly $\alpha$  forest to cosmological observations.

The future survey 4MOST [44] will be mounted on the Visible and Infrared Survey Telescope for Astronomy (VISTA) in Chile and will focus on the southern hemisphere with 812 fibers for high-resolution (HR in Tab. 2.1) and 1624 fibers for moderate-resolution (MR) observations.

### 2.3.2 Lyman- $\alpha$ forest based measurements

The use of the Ly $\alpha$  forest correlations to trace the IGM was first considered by Arthur Wolfe [47] in 1986. By the end of the 1990s, the development of the first cosmological simulations established the Ly $\alpha$  forest as a powerful probe of the matter [25]. Since then, the Ly $\alpha$  forest has become a major tool of modern cosmology, and a wide range of methods detailed here was put in place to extract maximum cosmological constraints.

#### Baryon acoustic oscillation measurements

The measurement of the Ly $\alpha$  forest auto-correlation function (Ly $\alpha$  $\times$ Ly $\alpha$ ), as well as its cross-correlation with other tracers given by Eqn. 2.24 is used to detect the BAO signal presented in Sec. 1.2.1.2.

The determination of BAO peak position in the correlations function provides a standard scale used to constrain  $\Lambda$ CDM parameters. The Ly $\alpha$  forest is currently the only cosmological probe able to detect BAO at redshift  $z > 2$ . It is mainly due to the extensive statistics provided by a single line-of-sight compared to one point for Boolean tracers. For a BAO measurement, it is necessary to use many spectra because the correlation signals are very low at separation near the BAO scale ( $r \sim 150$  Mpc).

The measurement of the Ly $\alpha$  forest auto-correlation was done for the first time by [48] with only 12 QSOs. Since then, the number of spectra used has only increased [49, 50, 51, 52, 53]. The drastic increase of Ly $\alpha$  forest observed with the BOSS surveys provided a measurement in [54] and a first detection of the BAO signal in [55] with Ly $\alpha$  $\times$ Ly $\alpha$  correlation. In parallel, the cross-correlation between the Ly $\alpha$  forest and the QSOs of the other lines-of-sight (Ly $\alpha$  $\times$ QSO) was measured in [56]. This cross-correlation can be combined with the Ly $\alpha$  $\times$ Ly $\alpha$  correlation in order to improve cosmological constraints [57, 58, 59]. The most recent measurement of Ly $\alpha$  $\times$ Ly $\alpha$  and Ly $\alpha$  $\times$ QSO correlations is given in [34] and represented in Fig. 2.7.

From those measurements, the BAO peak permits to estimate the ratios  $D_H/r_d$  and  $D_M/r_d$  where  $D_H$  and  $D_M$  are respectively the Hubble and comoving angular distance defined in Sec. 1.1.3. The comparison of these ratios with those of a fiducial cosmological model provides constraints on  $\Omega_{m,0}$  and  $H_0$  detailed in [60].

Other cross-correlation studies are performed to understand the Ly $\alpha$  systematics better or to use the full profile of the correlation function. In particular, the cross-correlation between Ly $\alpha$  forest and DLAs was also measured [61, 62] to characterize DLA contaminations. In Chap. 7 of this thesis, I compute the cross-correlation between the Ly $\alpha$  forest and the cosmic voids to detect matter velocity flows around voids at redshift  $z > 2$  [63].

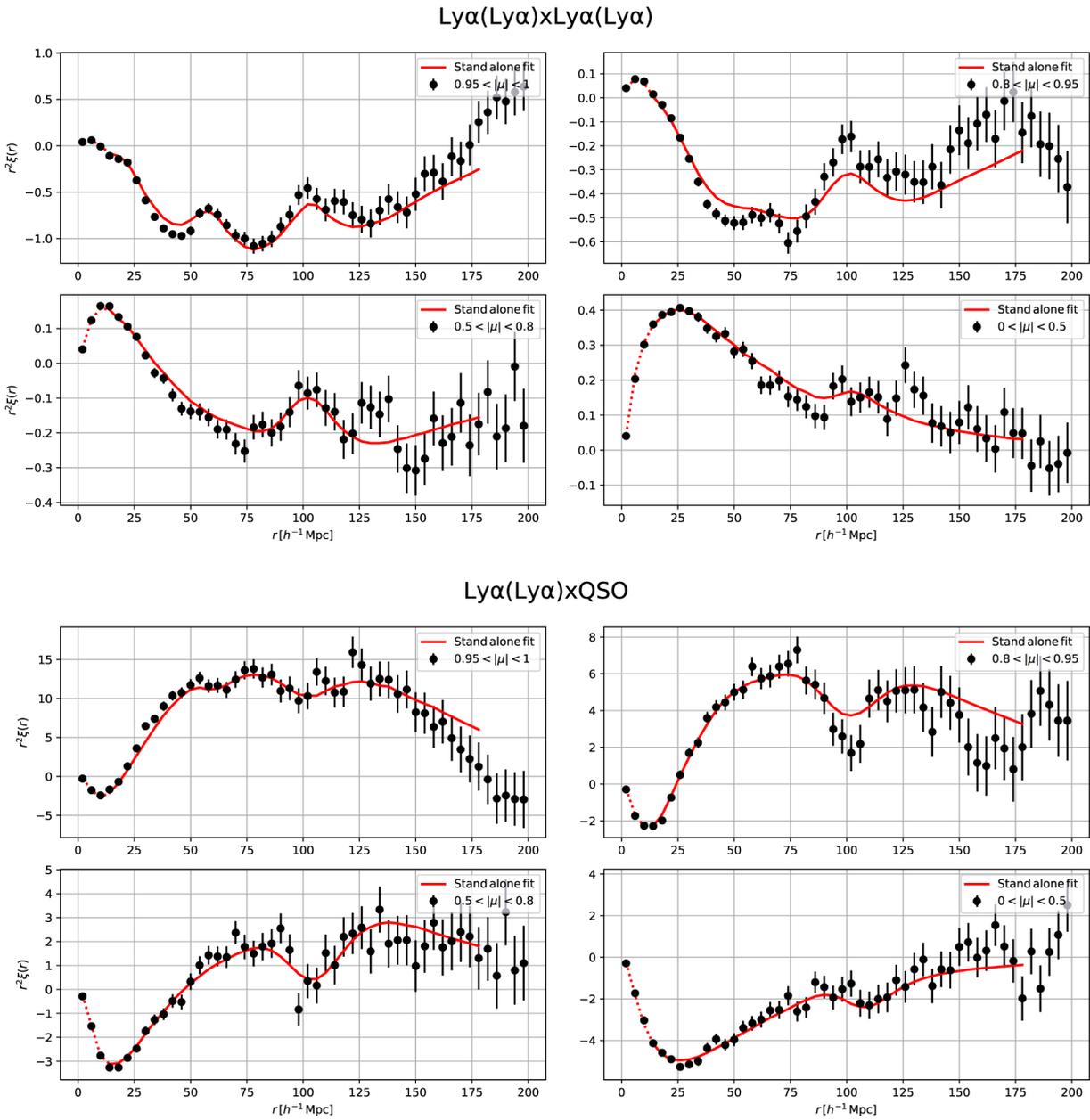


FIGURE 2.7 – (top) Ly $\alpha$  auto-correlation function and (bottom) Ly $\alpha$ -quasar cross-correlation measured with DR16 eBOSS data. The correlations are multiplied by the separation  $r^2$  to highlight the BAO scale. The black points are the measured correlation and the red curves give the best fit models, in four  $\mu$  bins. Extra knees or peaks caused by metal absorptions can be observed along the line-of-sight ( $|\mu| \in [0.95, 1]$ ) at separations  $\sim 20$  and  $\sim 60 h^{-1} \cdot \text{Mpc}$ . Figure taken from [34].

### One-dimensional Ly $\alpha$ forest power spectrum

The correlations of the Ly $\alpha$  forest can be computed along the line-of-sight when measuring the one-dimensional Ly $\alpha$  forest power spectrum ( $P_{1D,\alpha}$ ) given by Eqn. 2.27. This measure is the simplest way to access to small-scale correlations ( $\sim h^{-1} \cdot \text{Mpc}$ ) through a reasonably straightforward analysis. It can be realized from the two types of observations presented in previous section. Moderate-resolution surveys provide many Ly $\alpha$  forest, and therefore very high-precision measurement of  $P_{1D,\alpha}$ , but do not resolve the smallest scales. Both noise and spectrograph resolution have a very strong impact on the measurement. High-resolution observations probe matter at

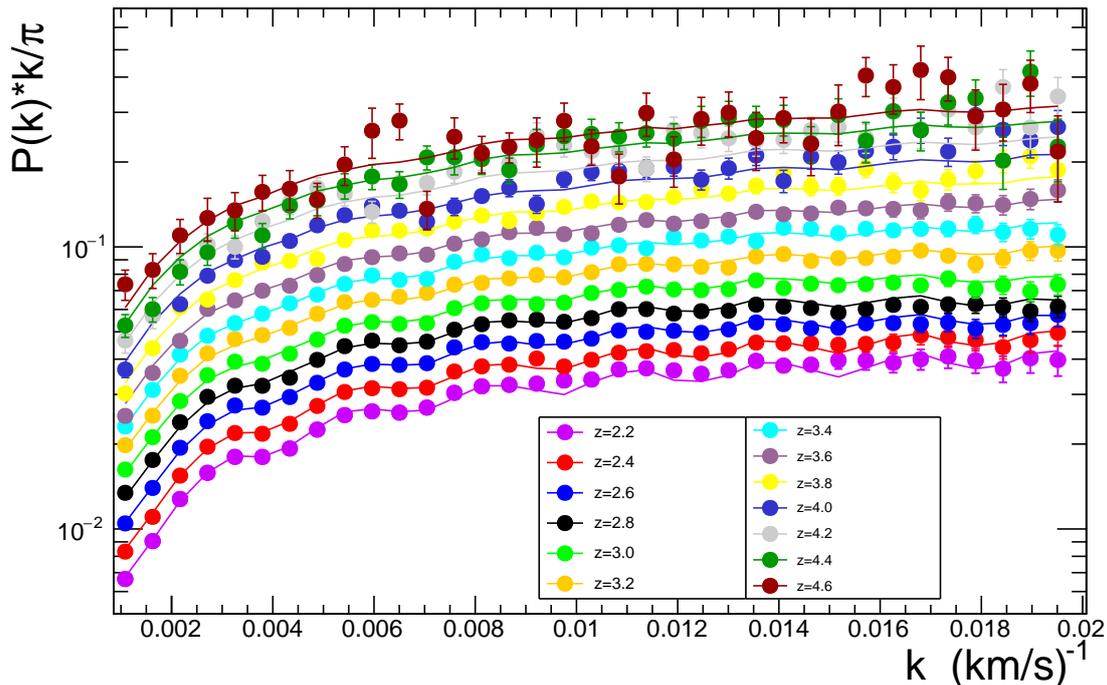


FIGURE 2.8 – One-dimensional Ly $\alpha$  forest power spectrum measurement with DR14 eBOSS data [64]. The error bars are a quadrature of statistics and systematics. The solid curves show the best-fit model. The oscillations arise from correlations between SiIII and Ly $\alpha$ , which occur at a rest-frame wavelength separation of 9.2  $\text{\AA}$ .

very small scales with high signal-to-noise ratio spectra. However, the statistics of these observations are relatively low. High- and moderate-resolution measurements are also impacted by the Ly $\alpha$  theoretical systematics detailed in Sec. 2.1.3.

The first measurement of  $P_{1D,\alpha}$  was performed on high-resolution data [24]. Subsequently, high-resolution observations coming from different surveys such as SQUAD, KODIAQ, or XQ-100 were used to improve this measurement [65, 66, 67, 68].

The  $P_{1D,\alpha}$  measurement with moderate-resolution spectrograph was first performed on a small sample of SDSS data in [69]. Subsequently, the increase of the Ly $\alpha$  forest statistic has largely improved this measurement with the BOSS [70] and eBOSS [64] surveys. In particular, the measurement performed on eBOSS, shown in Fig. 2.8, constitutes the most accurate measurement currently available and allows to reach high redshift bins. In this thesis, I will present in Chap. 5 a first measurement of  $P_{1D,\alpha}$  with DESI data.

The one-dimensional power spectrum is sensitive to the amplitude and slope of the matter power spectrum at redshift  $z > 2$ . The impact of cosmological parameters on  $P_{1D,\alpha}$  can be predicted using hydrodynamical simulations. Fitting data measurements with those predictions provides constraints on cosmological parameters such as  $\sigma_8$ ,  $n_s$ , and  $\Omega_m$ , but also on the IGM properties. In Chap. 3, I will present the simulations used to predict  $P_{1D,\alpha}$  with an accuracy adapted to the expected DESI measurement.

As shown in Sec. 1.3.4.1, the matter power spectrum is damped at small scales by a background of massive neutrinos. Due to its sensitivity to the amplitude of matter fluctuations at small scales,  $P_{1D,\alpha}$  is an ideal probe to constrain the sum of neutrino masses. Constraints were

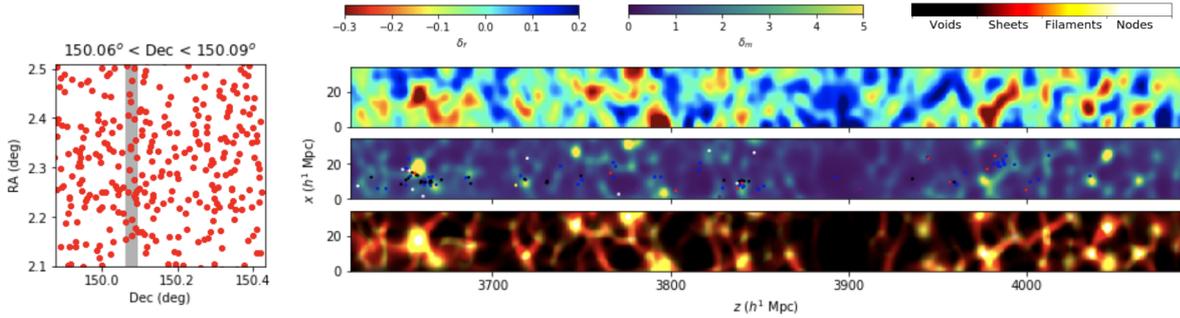


FIGURE 2.9 – Slice of the tomographic map taken from [71] using a Wiener filter on CLAMATO data (top right). Below the reconstruction, the inferred dark matter distribution from the TARDIS algorithm (middle right) and the inferred cosmic web classification (bottom right) are shown. The objects used for the reconstruction are pictured in red and the slice in gray in the left panel. The galaxies overlapping within the slice are represented by dots in the middle right panel.

obtained by coupling the one-dimensional power spectrum measurement with hydrodynamical simulations which include neutrinos [72, 73, 74], and combining  $P_{1D,\alpha}$  with the CMB, a cosmological probe sensitive to large scales. Thus, by coupling the  $P_{1D,\alpha}$  measurement from eBOSS given in Fig. 2.8 with the Planck data, the last measurement in [75] provided the best indirect constraint of the neutrino mass at the time.

As presented in Sec. 1.3.4.2, in the framework of warm dark matter models, the matter power spectrum undergoes a significant cutoff at small scales. Several studies [74, 75, 76, 77, 78, 79, 80] combined high- and moderate-resolution  $P_{1D,\alpha}$  measurements to obtain constraints on the warm dark matter mass. Other exotic dark matter models such as fuzzy dark matter predict a similar impact to WDM on the matter power spectrum. The one-dimensional power spectrum measurement was also used to constrain these models [81, 82].

### Ly $\alpha$ tomography

Ly $\alpha$  tomography consists of creating a three-dimensional contrast map by interpolating one-dimensional lines-of-sight. It is a method that requires a high density of lines-of-sight. The first idea of such a reconstruction was proposed in [83] and theorized in [84]. High- and moderate-resolution observations can be used to perform Ly $\alpha$  tomography. However, to obtain a 3D tomographic map with a high transverse spatial resolution, high-density observations are needed. Moderate-resolution observations are more suitable because they recover fainter targets with a shorter exposure time.

Ly $\alpha$  tomography was performed for the first time on data by the CLAMATO collaboration on a portion of the COSMOS field [71, 85]. The reconstructed map, shown in Fig. 2.9, uses the Ly $\alpha$  forest of QSOs and Lyman-break galaxies on a footprint of about  $0.2 \text{ deg}^2$ . This type of Ly $\alpha$  tomography is envisioned as a way to probe the cosmic web structure on a small volume. In particular, the ultimate goal of this type of survey is to obtain a transverse separation small enough to probe filaments ( $\sim h^{-1} \cdot \text{Mpc}$  size). A new dedicated survey called LATIS [86] is currently collecting data to create a Ly $\alpha$  tomographic map with the Inamori-Magellan Areal Camera and Spectrograph (IMACS). With future surveys like the ELT [87], it should be possible to produce  $h^{-1} \cdot \text{Mpc}$ -size tomographic maps and reconstruct filaments.

In this thesis, I present in Chap. 6 the first application of the Ly $\alpha$  tomography technique on a large field, using a dense and homogeneous footprint of eBOSS [88]. This study aims to

reconstruct a map that probes the large scale fluctuations of the matter distribution, over a large volume.

Tomographic map-making can be improved by using more developed interpolation techniques such as forward modeling [89, 90]. These methods will allow obtaining cosmological constraints while reconstructing the underlying matter density field.

### Other applications

Ly $\alpha$  forest can be used in other studies, such as the measurement of the Doppler width distribution expressed by Eqn. 2.10. As detailed in the recent study [91], this measurement is used to constrain reionization models and the Universe's thermal history.

On very high-resolution observations, the precise measurement of DLA properties is used to estimate the fine-structure constant at redshift  $z > 2$  [92]. Some cosmological models predict a redshift dependence for this constant and can be constrained with this kind of study.

## Bibliography

- [1] A. A. Meiksin, *The Physics of the Intergalactic Medium*, *Reviews of Modern Physics* **81** (2009) 1405.
- [2] M. McQuinn, *The Evolution of the Intergalactic Medium*, *Annual Review of Astronomy and Astrophysics* **54** (2016) 313.
- [3] B. T. Draine, *Physics of the interstellar and intergalactic medium*, Princeton series in astrophysics. Princeton University Press, Princeton, N.J, 2011.
- [4] M. Dijkstra, *Physics of Lyman Alpha Radiative Transfer*, *arXiv:1704.03416 [astro-ph]* (2017) .
- [5] J. Baur, *Determining the mass of cosmological neutrinos using Lyman-alpha forests*, phdthesis, Université Paris Saclay (COmUE), Sept., 2017.
- [6] P. Zarrouk, *Analyse des corrélations spatiales des quasars et implications cosmologiques avec le multi-spectrographe SDSS-IV eBOSS*, phdthesis, Université Paris-Saclay, Oct., 2018.
- [7] S. Chabanier, *Neutrinos and dark matter cosmology with the Lyman- $\alpha$  forest : the interplay between large-scale evolution and small-scale baryonic physics*, phdthesis, Université Paris-Saclay, Oct., 2020.
- [8] T. Etourneau, *Les forêts Lyman alpha du relevé eBOSS : comprendre les fonctions de corrélation et les systématiques*, phdthesis, Université Paris-Saclay, Sept., 2020.
- [9] T. A. Matthews and A. R. Sandage, *Optical Identification of 3C 48, 3C 196, and 3C 286 with Stellar Objects.*, *The Astrophysical Journal* **138** (1963) 30.
- [10] J. L. Greenstein and M. Schmidt, *The Quasi-Stellar Radio Sources 3C 48 and 3C 273.*, *The Astrophysical Journal* **140** (1964) 1.
- [11] J. Thorne, A. Robotham, L. Davies and S. Bellstedt, *AGN Unification Diagram*, Mar., 2022. 10.5281/zenodo.6381013.
- [12] E. E. Salpeter, *Accretion of Interstellar Matter by Massive Objects.*, *The Astrophysical Journal* **140** (1964) 796.
- [13] D. Lynden-Bell, *Galactic Nuclei as Collapsed Old Quasars*, *Nature* **223** (1969) 690.
- [14] R. R. J. Antonucci and J. S. Miller, *Spectropolarimetry and the nature of NGC 1068.*, *The Astrophysical Journal* **297** (1985) 621.
- [15] C. M. Urry, *AGN Unification: An Update*, *arXiv:astro-ph/0312545* (2003) .
- [16] F. Wang, J. Yang, X. Fan, J. F. Hennawi, A. J. Barth, E. Banados et al., *A Luminous Quasar at Redshift 7.642*, *The Astrophysical Journal Letters* **907** (2021) L1.
- [17] P. Laurent, S. Eftekharzadeh, J.-M. Le Goff, A. Myers, E. Burtin, M. White et al., *Clustering of quasars in SDSS-IV eBOSS: study of potential systematics and bias determination*, *Journal of Cosmology and Astroparticle Physics* **2017** (2017) 017.

- [18] N. Katz, D. H. Weinberg and L. Hernquist, *Cosmological Simulations with TreeSPH*, *The Astrophysical Journal Supplement Series* **105** (1996) 19.
- [19] M. Walther, J. Oñorbe, J. F. Hennawi and Z. Lukić, *New Constraints on IGM Thermal Evolution from the Ly $\alpha$  Forest Power Spectrum*, *The Astrophysical Journal* **872** (2019) 13.
- [20] W. L. W. Sargent, P. J. Young, A. Boksenberg and D. Tytler, *The distribution of Lyman-alpha absorption lines in the spectra of six QSOs: evidence for an intergalactic origin.*, *The Astrophysical Journal Supplement Series* **42** (1980) 41.
- [21] J. E. Gunn and B. A. Peterson, *On the density of neutral hydrogen intergalactic space*, *American Astronomical Society* (1965) .
- [22] D. H. Weinberg, J. Miralda-Escude, L. Hernquist and N. Katz, *A Lower Bound on the Cosmic Baryon Density*, *The Astrophysical Journal* **490** (1997) 564.
- [23] D. H. Weinberg, L. Hernquist, N. Katz, R. Croft and J. Miralda-Escudé, *Hubble Flow Broadening of the LY $\alpha$  Forest and its Implications*. eprint: arXiv:astro-ph/9709303, Jan., 1997.
- [24] R. A. C. Croft, D. H. Weinberg, N. Katz and L. Hernquist, *Recovery of the Power Spectrum of Mass Fluctuations from Observations of the Lyman-alpha Forest*, *The Astrophysical Journal* **495** (1998) 44.
- [25] R. Croft, D. Weinberg, N. Katz and L. Hernquist, *Cosmology from the structure of the Lyman-alpha forest*, *arXiv:astro-ph/9801255* (1998) .
- [26] D. H. Weinberg, R. Dav'e, N. Katz and J. A. Kollmeier, *The Lyman-alpha Forest as a Cosmological Tool*, *AIP Conference Proceedings* **666** (2003) 157.
- [27] A. M. Wolfe, E. Gawiser and J. X. Prochaska, *Damped Lyman alpha Systems*, *Annual Review of Astronomy and Astrophysics* **43** (2005) 861.
- [28] J. X. Prochaska, J. M. O'Meara, S. Herbert-Fort, S. Burles, G. E. Prochter and R. A. Bernstein, *Super-Solar Super Lyman Limit Systems*, *The Astrophysical Journal* **648** (2006) L97.
- [29] R. J. Weymann, S. L. Morris, C. B. Foltz and P. C. Hewett, *Comparisons of the Emission-Line and Continuum Properties of Broad Absorption Line and Normal Quasi-stellar Objects*, *The Astrophysical Journal* **373** (1991) 23.
- [30] R. Weymann, *BAL QSOs: Properties and Problems - an Optical Spectroscopist's Overview*, *Mass Ejection from Active Galactic Nuclei; ASP Conference Series* **128** (1997) 3.
- [31] L. Ennesser, P. Martini, A. Font-Ribera and I. Pérez-Ràfols, *The impact and mitigation of broad absorption line quasars in Lyman- $\alpha$  forest correlations*, *arXiv:2111.09439 [astro-ph]* (2021) .
- [32] F. Calura, E. Tescari, V. D'Odorico, M. Viel, S. Cristiani, T.-S. Kim et al., *The Lyman alpha forest flux probability distribution at  $z > 3$* , *Monthly Notices of the Royal Astronomical Society* **422** (2012) 3019.

- [33] U. Seljak, *Bias, redshift space distortions and primordial nongaussianity of nonlinear transformations: application to Lyman alpha forest*, *Journal of Cosmology and Astroparticle Physics* **2012** (2012) 004.
- [34] H. d. M. d. Bourboux, J. Rich, A. Font-Ribera, V. d. S. Agathe, J. Farr, T. Etourneau et al., *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Baryon acoustic oscillations with Lyman- $\alpha$  forests*, *The Astrophysical Journal* **901** (2020) 153.
- [35] J. M. O'Meara, N. Lehner, J. C. Howk, J. X. Prochaska, A. J. Fox, M. A. Swain et al., *THE FIRST DATA RELEASE OF THE KODIAQ SURVEY*, *The Astronomical Journal* **150** (2015) 111.
- [36] M. T. Murphy, G. G. Kacprzak, G. A. D. Savorgnan and R. F. Carswell, *The UVES Spectral Quasar Absorption Database (SQUAD) Data Release 1: The first 10 million seconds*, *arXiv:1810.06136 [astro-ph]* (2018) .
- [37] S. Lopez, V. D'Odorico, S. L. Ellison, G. D. Becker, L. Christensen, G. Cupani et al., *XQ-100: A legacy survey of one hundred  $3.5 < z < 4.5$  quasars observed with VLT/XSHOOTER*, *Astronomy & Astrophysics* **594** (2016) A91.
- [38] J. M. O'Meara, *The Second Data Release of the KODIAQ Survey*, *The Astronomical Journal* (2017) 5.
- [39] C. W. Danforth, B. A. Keeney, E. M. Tilton, J. M. Shull, J. T. Stocke, M. Stevans et al., *An HST/COS Survey of the Low-redshift Intergalactic Medium. I. Survey, Methodology, and Overall Results*, *The Astrophysical Journal* **817** (2016) 111.
- [40] K. S. Dawson, J.-P. Kneib, W. J. Percival, S. Alam, F. D. Albareti, S. F. Anderson et al., *The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data*, *The Astronomical Journal* **151** (2016) 44.
- [41] DESI collaboration, *The DESI Experiment Part I: Science, Targeting, and Survey Design*, *arXiv:1611.00036 [astro-ph]* (2016) .
- [42] DESI collaboration, *The DESI Experiment Part II: Instrument Design*, *arXiv:1611.00037 [astro-ph]* (2016) .
- [43] M. M. Pieri, S. Bonoli, J. Chaves-Montero, I. Paris, M. Fumagalli, J. S. Bolton et al., *WEAVE-QSO: A Massive Intergalactic Medium Survey for the William Herschel Telescope*, *arXiv:1611.09388 [astro-ph]* (2016) .
- [44] R. S. de Jong, O. Agertz, A. A. Berbel, J. Aird, D. A. Alexander, A. Amarsi et al., *4MOST: Project overview and information for the First Call for Proposals*, *Published in The Messenger vol. 175 pp. 3-11* (2019) 9 pages.
- [45] K. Abazajian, *The Second Data Release of the Sloan Digital Sky Survey*, *The Astronomical Journal* **128** (2004) 502.
- [46] D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. A. Prieto, S. F. Anderson et al., *SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way Galaxy, and Extra-Solar Planetary Systems*, *The Astronomical Journal* **142** (2011) 72.

- [47] A. M. Wolfe, J. D. Barrow, P. J. E. Peebles and D. W. Sciama, *New evidence from the Lyman-alpha forest concerning the formation of galaxies*, *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* **320** (1986) 503.
- [48] L. Zuo and J. R. Bond, *The Transmission Correlation in the QSO LY alpha Forest Produced by Finite Width Lines*, *The Astrophysical Journal* **423** (1994) 73.
- [49] J. Liske, J. K. Webb, G. M. Williger, A. Fernández-Soto and R. F. Carswell, *Large-scale structure in the Lyman- $\alpha$  forest—II. Analysis of a group of 10 QSOs*, *Monthly Notices of the Royal Astronomical Society* **311** (2000) 657.
- [50] P. McDonald, J. Miralda-Escude, M. Rauch, W. L. W. Sargent, T. A. Barlow, R. Cen et al., *The Observed Probability Distribution Function, Power Spectrum, and Correlation Function of the Transmitted Flux in the Ly $\alpha$  Forest*, *The Astrophysical Journal* **543** (2000) 1.
- [51] R. A. C. Croft, D. H. Weinberg, M. Bolte, S. Burles, L. Hernquist, N. Katz et al., *Towards a Precise Measurement of Matter Clustering: Lyman-alpha Forest Data at Redshifts 2-4*, *The Astrophysical Journal* **581** (2002) 20.
- [52] G. D. Becker, W. L. W. Sargent and M. Rauch, *Large-Scale Correlations in the Ly $\alpha$  Forest at  $z = 3-4^*$* , *The Astrophysical Journal* **613** (2004) 61.
- [53] D. Kirkman, D. Tytler, D. Lubin and J. Charlton, *Continuous statistics of the Ly $\alpha$  forest at  $0 < z < 1.6$ : the mean flux, flux distribution and autocorrelation from HST FOS spectra*, *Monthly Notices of the Royal Astronomical Society* **376** (2007) 1227.
- [54] A. Slosar, A. Font-Ribera, M. M. Pieri, J. Rich, J.-M. L. Goff, É. Aubourg et al., *The Lyman-alpha forest in three dimensions: measurements of large scale flux correlations from BOSS 1st-year data*, *Journal of Cosmology and Astroparticle Physics* **2011** (2011) 001.
- [55] N. G. Busca, T. Delubac, J. Rich, S. Bailey, A. Font-Ribera, D. Kirkby et al., *Baryon Acoustic Oscillations in the Ly- $\alpha$  forest of BOSS quasars*, *Astronomy & Astrophysics* **552** (2013) A96.
- [56] A. Font-Ribera, D. Kirkby, N. Busca, J. Miralda-Escudé, N. P. Ross, A. Slosar et al., *Quasar-Lyman  $\alpha$  Forest Cross-Correlation from BOSS DR11 : Baryon Acoustic Oscillations*, *Journal of Cosmology and Astroparticle Physics* **2014** (2014) 027.
- [57] T. Delubac, J. E. Bautista, N. G. Busca, J. Rich, D. Kirkby, S. Bailey et al., *Baryon Acoustic Oscillations in the Ly $\alpha$  forest of BOSS DR11 quasars*, *Astronomy & Astrophysics* **574** (2015) A59.
- [58] J. E. Bautista, N. G. Busca, J. Guy, J. Rich, M. Blomqvist, H. d. M. d. Bourboux et al., *Measurement of BAO correlations at  $z=2.3$  with SDSS DR12 Ly $\alpha$ -Forests*, *Astronomy & Astrophysics* **603** (2017) A12.
- [59] H. d. M. d. Bourboux, J.-M. L. Goff, M. Blomqvist, N. G. Busca, J. Guy, J. Rich et al., *Baryon acoustic oscillations from the complete SDSS-III Ly $\alpha$ -quasar cross-correlation function at  $z=2.4$* , *Astronomy & Astrophysics* **608** (2017) A130.

- [60] EBOSS collaboration, *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological Implications from two Decades of Spectroscopic Surveys at the Apache Point observatory*, *Physical Review D* **103** (2021) 083533.
- [61] A. Font-Ribera, J. Miralda-Escudé, E. Arnau, B. Carithers, K.-G. Lee, P. Noterdaeme et al., *The large-scale cross-correlation of Damped Lyman Alpha Systems with the Lyman Alpha Forest: First Measurements from BOSS*, *Journal of Cosmology and Astroparticle Physics* **2012** (2012) 059.
- [62] I. Pérez-Ràfols, A. Font-Ribera, J. Miralda-Escudé, M. Blomqvist, S. Bird, N. Busca et al., *The SDSS-DR12 large-scale cross-correlation of Damped Lyman Alpha Systems with the Lyman Alpha Forest*, *Monthly Notices of the Royal Astronomical Society* **473** (2018) 3019.
- [63] C. Ravoux, E. Armengaud, J. Bautista, J.-M. L. Goff, N. Palanque-Delabrouille, J. Rich et al., *First measurement of the correlation between cosmic voids and the Lyman- $\alpha$  forest*, *arXiv:2203.11045 [astro-ph]* (2022) .
- [64] S. Chabanier, N. Palanque-Delabrouille, C. Yèche, J.-M. L. Goff, E. Armengaud, J. Bautista et al., *The one-dimensional power spectrum from the SDSS DR14 Ly $\alpha$  forests*, *Journal of Cosmology and Astroparticle Physics* **2019** (2019) 017.
- [65] T.-S. Kim, M. Viel, M. G. Haehnelt, R. F. Carswell and S. Cristiani, *The power spectrum of the flux distribution in the Lyman-alpha forest of a Large sample of UVES QSO Absorption Spectra (LUQAS)*, *Monthly Notices of the Royal Astronomical Society* **347** (2004) 355.
- [66] V. Iršič, M. Viel, T. A. M. Berg, V. D’Odorico, M. G. Haehnelt, S. Cristiani et al., *The Lyman-alpha forest power spectrum from the XQ-100 Legacy Survey*, *Monthly Notices of the Royal Astronomical Society* (2016) stw3372.
- [67] M. Walther, J. F. Hennawi, H. Hiss, J. Oñorbe, K.-G. Lee, A. Rorai et al., *A New Precision Measurement of the Small-scale Line-of-sight Power Spectrum of the Ly $\alpha$  Forest*, *The Astrophysical Journal* **852** (2018) 22.
- [68] N. G. Karaçaylı, N. Padmanabhan, A. Font-Ribera, V. Iršič, M. Walther, D. Brooks et al., *Optimal 1D Ly  $\alpha$  forest power spectrum estimation – II. KODIAQ, SQUAD, and XQ-100*, *Monthly Notices of the Royal Astronomical Society* **509** (2022) 2842.
- [69] P. McDonald, U. Seljak, S. Burles, D. J. Schlegel, D. H. Weinberg, D. Shih et al., *The Lyman-alpha Forest Power Spectrum from the Sloan Digital Sky Survey*, *The Astrophysical Journal Supplement Series* **163** (2006) 80.
- [70] N. Palanque-Delabrouille, C. Yèche, A. Borde, J.-M. L. Goff, G. Rossi, M. Viel et al., *The one-dimensional Ly-alpha forest power spectrum from BOSS*, *Astronomy & Astrophysics* **559** (2013) A85.
- [71] B. Horowitz, K.-G. Lee, M. Ata, T. Müller, A. Krolewski, J. X. Prochaska et al., *Second Data Release of the COSMOS Lyman-alpha Mapping and Tomographic Observation: The First 3D Maps of the Large-Scale Cosmic Web at  $2.05 < z < 2.55$* , *arXiv:2109.09660 [astro-ph]* (2021) .

- [72] U. Seljak, A. Slosar and P. McDonald, *Cosmological parameters from combining the Lyman-alpha forest with CMB, galaxy clustering and SN constraints*, *Journal of Cosmology and Astroparticle Physics* **2006** (2006) 014.
- [73] N. Palanque-Delabrouille, C. Yèche, J. Baur, C. Magneville, G. Rossi, J. Lesgourgues et al., *Neutrino masses and cosmology with Lyman-alpha forest power spectrum*, *Journal of Cosmology and Astroparticle Physics* **2015** (2015) 011.
- [74] C. Yeche, N. Palanque-Delabrouille, J. Baur and H. d. M. d. BourBoux, *Constraints on neutrino masses from Lyman-alpha forest power spectrum with BOSS and XQ-100*, *Journal of Cosmology and Astroparticle Physics* **2017** (2017) 047.
- [75] N. Palanque-Delabrouille, C. Yèche, N. Schöneberg, J. Lesgourgues, M. Walther, S. Chabanier et al., *Hints, neutrino bounds and WDM constraints from SDSS DR14 Lyman- $\alpha$  and Planck full-survey data*, *Journal of Cosmology and Astroparticle Physics* **2020** (2020) 038.
- [76] M. Viel, J. Lesgourgues, M. G. Haehnelt, S. Matarrese and A. Riotto, *Constraining Warm Dark Matter candidates including sterile neutrinos and light gravitinos with WMAP and the Lyman-alpha forest*, *Physical Review D* **71** (2005) 063534.
- [77] M. Viel, G. D. Becker, J. S. Bolton, M. G. Haehnelt, M. Rauch and W. L. W. Sargent, *How cold is cold dark matter? Small scales constraints from the flux power spectrum of the high-redshift Lyman-alpha forest*, *Physical Review Letters* **100** (2008) 041304.
- [78] M. Viel, G. D. Becker, J. S. Bolton and M. G. Haehnelt, *Warm Dark Matter as a solution to the small scale crisis: new constraints from high redshift Lyman-alpha forest data*, *Physical Review D* **88** (2013) 043502.
- [79] J. Baur, N. Palanque-Delabrouille, C. Yèche, C. Magneville and M. Viel, *Lyman-alpha Forests cool Warm Dark Matter*, *Journal of Cosmology and Astroparticle Physics* **2016** (2016) 012.
- [80] J. Baur, N. Palanque-Delabrouille, C. Yèche, A. Boyarsky, O. Ruchayskiy, É. Armengaud et al., *Constraints from Ly- $\alpha$  forests on non-thermal dark matter including resonantly-produced sterile neutrinos*, *Journal of Cosmology and Astroparticle Physics* **2017** (2017) 013.
- [81] V. Iršič, M. Viel, M. G. Haehnelt, J. S. Bolton and G. D. Becker, *First constraints on fuzzy dark matter from Lyman- $\alpha$  forest data and hydrodynamical simulations*, *Physical Review Letters* **119** (2017) 031302.
- [82] E. Armengaud, N. Palanque-Delabrouille, C. Yèche, D. J. E. Marsh and J. Baur, *Constraining the mass of light bosonic dark matter using SDSS Lyman- $\alpha$  forest*, *Monthly Notices of the Royal Astronomical Society* **471** (2017) 4606.
- [83] P. Petitjean, *Clustering of absorption line systems*, *arXiv:astro-ph/9608115* (1996) .
- [84] C. Pichon, J. Vergely, E. Rollinde, S. Colombi and P. Petitjean, *Inversion of the Lyman  $\alpha$  forest: three-dimensional investigation of the intergalactic medium*, *Monthly Notices of the Royal Astronomical Society* **326** (2001) 597.

- [85] K.-G. Lee, A. Krolewski, M. White, D. Schlegel, P. E. Nugent, J. F. Hennawi et al., *First Data Release of the COSMOS Ly  $\alpha$  Mapping and Tomography Observations: 3D Ly  $\alpha$  Forest Tomography at  $2.05 < z < 2.55$* , *The Astrophysical Journal Supplement Series* **237** (2018) 31.
- [86] A. B. Newman, G. C. Rudie, G. A. Blanc, D. D. Kelson, S. Rhoades, T. Hare et al., *LATIS: The Ly $\alpha$  Tomography IMACS Survey*, *The Astrophysical Journal* **891** (2020) 147.
- [87] J. Japelj, C. Laigle, M. Puech, C. Pichon, H. Rahmani, Y. Dubois et al., *Simulating MOS science on the ELT: Ly $\alpha$  forest tomography*, *Astronomy & Astrophysics* **632** (2019) A94.
- [88] C. Ravoux, E. Armengaud, M. Walther, T. Etourneau, D. Pomarède, N. Palanque-Delabrouille et al., *A tomographic map of the large-scale matter distribution using the eBOSS Stripe 82 Ly- $\alpha$  forest*, *Journal of Cosmology and Astroparticle Physics* **2020** (2020) 010.
- [89] B. Horowitz, K.-G. Lee, M. White, A. Krolewski and M. Ata, *TARDIS Paper I: A Constrained Reconstruction Approach to Modeling the  $z \sim 2.5$  Cosmic Web Probed by Lyman-alpha Forest Tomography*, *The Astrophysical Journal* **887** (2019) 61.
- [90] N. Porqueres, O. Hahn, J. Jasche and G. Lavaux, *A hierarchical field-level inference approach to reconstruction from sparse Lyman- $\alpha$  forest data*, *arXiv:2005.12928 [astro-ph]* (2020) .
- [91] H. Hiss, M. Walther, J. Oñorbe and J. F. Hennawi, *A Novel Statistical Method for Measuring the Temperature-Density Relation in the IGM Using the b-NHI Distribution of Absorbers in the Ly $\alpha$  Forest*, *The Astrophysical Journal* **876** (2019) 71.
- [92] D. Milaković, C.-C. Lee, R. F. Carswell, J. K. Webb, P. Molaro and L. Pasquini, *A new era of fine structure constant measurements at high redshift*, *Monthly Notices of the Royal Astronomical Society* **500** (2020) 1.

# 3

---

## Simulating the Lyman- $\alpha$ forest

---

*“The Machines are not super-brains in Sunday supplement sense, - although they are so pictured in the Sunday supplements. It is merely that in their own particular province of collecting and analyzing a nearly infinite number of data and relationships thereof, in nearly infinitesimal time, they have progressed beyond the possibility of detailed human control.”*

– Isaac Asimov, *I, Robot - The Evitable Conflict*, 1950

**Contents**

---

<b>3.1</b>	<b>Log-normal mocks for Lyman-<math>\alpha</math> forest</b>	<b>96</b>
3.1.1	Saclay mocks	96
3.1.2	DESI-Lite mocks	99
3.1.3	Modeling systematics in mocks	100
<b>3.2</b>	<b>Cosmological simulations</b>	<b>102</b>
3.2.1	Overview	102
3.2.2	Simulation pipeline	106
3.2.3	Simulating the one-dimensional Lyman- $\alpha$ forest power spectrum	108
3.2.4	The three-dimensional power spectrum of the Lyman- $\alpha$ forest	112
	<b>Bibliography</b>	<b>120</b>

---

SINCE cosmology is concerned with determining the Universe’s properties, researchers have no choice but to work in a complex laboratory. Indeed, only a minor portion of the Universe is accessible, and contrarily to experiments performed on Earth, it is impossible to modify the experimental conditions. In this context, numerical simulations are necessary to perform tests or constrain cosmological models from observational data.

More precisely, simulations are essential when analytical models become ineffective. In Chap. 1, a global model of the Universe, followed by the linear perturbations theory, was derived. Some approaches such as non-linear perturbation theory or effective field theory overcome the linear approximation. These theories are helpful to model smaller scales or smaller redshifts. However, these corrections usually involve free parameters which must be calibrated. Furthermore, they are generally not accurate for all scales. For instance, it is impossible to model analytically the smallest scales of the Ly $\alpha$  forest detailed in Chap. 2. In addition, for the Ly $\alpha$  forest, it is imperative to model the properties of baryonic matter which involve complex physics.

In this framework, the non-linear equations given in Sec. 1.3 must be solved numerically. However, solving exactly these equations is usually very expensive in terms of computing resources. Designing a numerical experiment always involves playing with a trade-off between accuracy and physics modeling on the one hand, and memory usage and computation time on the other hand. Therefore, it is necessary to carefully choose the numerical method employed according to the study and the scales considered. The complexity of a simulation is mainly determined by the cosmological fluids modeled and the numerical scheme used to solve the equations.

A first category of numerical experiments considered here is the creation of synthetic data, also called **mocks**. Their goal is to mimic a large amount of data, and the evolution of cosmological fluids is generally realized with approximate methods. Mocks are very close to data analysis and are designed to study the impact of instrumental, analytical, and theoretical systematics on the data. In Sec. 3.1, I present the log-normal mocks used in this thesis. These mocks model the Ly $\alpha$  forest and the impact of systematics on the one-dimensional power-spectrum in Chap. 5, tomography in Chap. 6, and cross-correlation with voids in Chap. 7.

However, approximate mocks are inadequate to provide accurate predictions and quantify the impact of cosmological parameters on the observables. In that case, exact methods implemented in **cosmological simulations** are needed. In practice, the distinction between mocks and cosmological simulations is blurred, and there are many numerical methods halfway between these two categories. A simulation models cosmological components as particles or fluids, and makes them evolve in time by modeling their interactions. There exists a wide range of methods, from large distance force models to full hydrodynamic resolution. In particular, I detail in Sec. 3.2 the hydrodynamical simulations used in this thesis.

This chapter is based on the review article [1] and previous manuscripts [2, 3]. I also rely on the article [4] which I co-authored. For the analysis in Sec. 3.2.4, I developed the `lyapower`<sup>1</sup> python package. All algorithms used in this section are documented in this public package.

---

1. <https://github.com/corentinravoux/lyapower/>

## 3.1 Log-normal mocks for Lyman- $\alpha$ forest

Log-normal mocks are based on approximate methods used to generate data quickly for a large volume and a reasonable resolution. Generating similar data sets with cosmological N-body or hydrodynamical simulations is unachievable considering the current supercomputer capacity. However, most log-normal mock parameters are calibrated using the results of this type of simulation.

A Ly $\alpha$  log-normal mock requires a method to generate a density field, and a model relating this field to the Ly $\alpha$  forest. In this section, the mocks use **Gaussian random fields (GRF)** to generate the matter density field. The Ly $\alpha$  forest is then constructed using the FGPA (see Sec. 2.1.2.2) and the **log-normal approximation** [5, 6]. In Sec. 3.1.1 and 3.1.2, I present two softwares used to generate the fraction of transmitted flux  $F$  as defined by Eqn. 2.5. The pipeline used to compute mock observations from these transmissions is described in Sec. 3.1.3.

### 3.1.1 Saclay mocks

I generated synthetic data with the **SaclayMocks**<sup>2</sup> software in order to interpret measurements of Chap. 6 and 7. This software assumes that cosmological fields such as matter density are Gaussian random fields. The **SaclayMocks** was developed to reproduce the observed statistical properties of the Ly $\alpha$  forest on relatively large scales [2, 7, 8]. They were used to test the impact of systematics on the eBOSS Ly $\alpha$  BAO analysis in [9]. These mocks are similar to **LyaCoLoRe** [10] which is an alternative implementation for systematics testing. Both mocks can reproduce the footprints of eBOSS or DESI surveys which will be detailed in Chap. 4.

#### Gaussian random field

Large-scale matter density fluctuations are generated using Gaussian random fields. Similarly to the derivation performed in Sec. 2.2.2, **SaclayMocks** software assumes that matter fluctuations can be separated into large and small scale components, respectively noted  $\delta_L$  and  $\delta_S$ . For the large-scale matter field, a GRF noted  $\delta_{\text{GRF}}$  is generated on a grid of constant voxel length  $d_{\text{cell}} = 2.19 h^{-1} \cdot \text{Mpc}$  (at  $z = 0$ ) along the three spatial directions. In order to have the  $\delta_L$  field, for a given power spectrum  $P(\vec{k})$ , the GRF field is Fourier transformed and multiplied by  $\sqrt{P(\vec{k})/d_{\text{cell}}^3}$ . The large-scale matter field  $\delta_L$  in real space is thus defined by

$$\delta_L(\vec{r}) = \mathcal{F}^{-1} \left[ \sqrt{\frac{P(\vec{k})}{d_{\text{cell}}^3}} \times \mathcal{F}[\delta_{\text{GRF}}(\vec{r})] \right]. \quad (3.1)$$

In order to include RSD in the Ly $\alpha$  forest, peculiar velocities  $\vec{u}$  and the associated peculiar velocity gradient fields must also be computed. The velocity field is calculated in Fourier space from  $\delta_L(\vec{k})$  using the linear approximation given by Eqn. 1.130. The velocity gradient is defined by

$$\eta_{ij} = -\frac{1}{\mathcal{H}} \frac{\partial u_i}{\partial x_j}. \quad (3.2)$$

2. <https://github.com/igmhub/SaclayMocks>

Using this equation and Eqn. 1.130, the velocity gradient is generated in Fourier space directly from the  $\delta_L(\vec{k})$  field. Three velocity boxes and six velocity gradient boxes along all directions are constructed at redshift  $z = 0$  in real space using a Fourier transform.

A QSO density field that reproduces the clustering observed in data should also be generated. This field is created from  $\delta_L(\vec{k})$  after multiplying it by the quasar power spectrum at the redshift of the Ly $\alpha$  forest. This field is defined for three different redshift bins  $z_i$ :

$$\delta_{q,i}(\vec{k}) = \delta_L(\vec{k}) \sqrt{\frac{P_{q,i}(k)}{d_{\text{cell}}^3}}, \quad (3.3)$$

where the quasar power spectra  $P_{q,i}$  is calculated from the linear matter power spectrum  $P_m$  at  $z = 0$  and the quasar bias  $b_q$  according to:

$$P_{q,i}(k) = \mathcal{F} \left[ \ln \left( 1 + \mathcal{F}^{-1} \left[ (b_q(z_i) G(z_i))^2 P_m(z = 0, k) \right] \right) \right]. \quad (3.4)$$

The  $G(z)$  terms corresponds to the growth function  $D_+(z)$  defined in Sec. 1.3.3.1, normalized such that  $G(z = 0) = 1$ . This function can be computed using a Boltzmann code.

By performing the inverse Fourier transform of the three  $\delta_{q,i}(\vec{k})$  boxes, the QSO density field is obtained in real space. Quasars are then generated from the following log-normal distribution  $\mathcal{P}_q$ :

$$\mathcal{P}_q \propto \exp \left( \frac{b_q(z)(1+z_0)}{b_q(z_0)(1+z)} \delta_{q,0}(\vec{r}) \right), \quad (3.5)$$

where  $z_0$  is the redshift of the box  $\delta_{q,0}(\vec{r})$  closest to the considered redshift  $z$ .

In conclusion, a box of large-scale matter fluctuations  $\delta_L$  at  $z = 0$ , six boxes of velocity gradient components at  $z = 0$ , and a QSO catalog are produced. The QSO catalog is adapted to match a given survey's geometry and redshift distribution. All boxes are in real space and hereafter the  $\vec{r}$  notation is dropped for clarity.

### FGPA model in mocks

The FGPA model in Eqn. 2.20 is used to relate the optical depth of the Ly $\alpha$  forest to the baryon contrast density such that  $\tau_\alpha \propto (1 + \delta_b)^\beta$ , with  $\beta = 2 - 0.7(\gamma - 1)$ .

Here, we adapt this model to be expressed in observational variables rather than IGM parameters. First, in the development of Eqn. 2.20, the peculiar velocity of the IGM gas was neglected. Similarly to Sec. 2.2.2, we introduce RSD with the conservation of optical depth between real and redshift space given by Eqn. 2.31 and 2.32:

$$\tau_\alpha^s = \left( 1 - \frac{1}{\mathcal{H}} \frac{\partial \vec{u} \cdot \hat{r}}{\partial r_\parallel} \right) \tau_\alpha = (1 + \eta_\parallel) \tau_\alpha, \quad (3.6)$$

where  $\eta_\parallel$  is the velocity gradient along the line-of-sight. Consequently, the FGPA model is modified such that:

$$\tau_\alpha^s \propto (1 + \delta_b)^\beta (1 + \eta_\parallel). \quad (3.7)$$

Since we consider large scales, we assume that the baryon density contrast  $\delta_b$  follows the large-scale matter one  $\delta_L$ . To generate the Ly $\alpha$  forest, the `SaclayMocks` software uses the log-normal approximation of the optical depth [5, 6] which models the matter distribution by a

log-normal random field. As  $\delta_L$  is a GRF normalized with a fixed power spectrum, the log-normal approximation expresses the optical depth such that

$$\tau_\alpha^s \propto \exp(\beta\delta_L + \eta_\parallel). \quad (3.8)$$

In order to reproduce the redshift dependence of Ly $\alpha$  observations, this relation is further parameterized. The fraction of transmitted flux  $F = \exp(-\tau_\alpha^s)$  given by Eqn. 2.5 is finally expressed by

$$F = \exp \left[ -a_{\text{GP}}(z) \exp \left( b_{\text{GP}}(z)G(z)(\delta_L + \delta_S + c_{\text{GP}}(z)\eta_\parallel) \right) \right]. \quad (3.9)$$

The Gunn-Peterson parameters  $a_{\text{GP}}(z)$ ,  $b_{\text{GP}}(z)$  and  $c_{\text{GP}}(z)$  are tuned to reproduce the mean transmitted flux fraction  $\bar{F}(z)$ , the Ly $\alpha$  bias  $b_\alpha(z)$  and the RSD parameter  $\beta_\alpha(z)$ . The  $c_{\text{GP}}(z)$  parameter controls RSD, and its value is actually close to  $\beta_\alpha(z)$ . For the mocks used in Chap. 6 and 7, the value used for this tuning are taken from [9]. In particular, the redshift dependence of the  $b_{\text{GP}}(z)$  parameter is entirely taken into account by the growth function. To obtain an IGM equation-of-state  $\gamma = 1.6$  [11],  $b_{\text{GP}}$  is set equal to  $\beta = 2 - 0.7(\gamma - 1) = 1.58$ . In Eqn. 3.9, the field  $\delta_S$  is added to mimic small-scale flux fluctuations that are missing due to the relatively large voxel size ( $2.19 h^{-1}\text{Mpc}$ ). It is tuned to match the observed one-dimensional Ly $\alpha$  power spectrum, which is the main topic of Chap. 5. A one-dimensional GRF is created to generate the  $\delta_S$  field:

$$\delta_S = \frac{\sigma_S(z)}{\sigma_S(z_{\text{eff}})} \mathcal{F}^{-1} \left[ \sqrt{\frac{P_S(z_{\text{eff}}, k)}{d_{\text{pix}}}} \times \mathcal{F}[\delta_{1\text{D}, \text{GRF}}(r)] \right], \quad (3.10)$$

where  $z_{\text{eff}}$  is the mean redshift of the modeled Ly $\alpha$  forest. Unlike the other GRFs,  $\delta_S$  is generated along the lines-of-sight. The size  $d_{\text{pix}}$  corresponds to the size of a line-of-sight pixel. The factor  $\sigma_S$  is the standard deviation of the field  $\delta_S$  and is used to renormalize the one-dimensional power spectrum at any redshift of the line-of-sight. Finally, the power spectrum  $P_S$  is tuned to reproduce the one-dimensional Ly $\alpha$  power spectrum measured in [12] at  $z_{\text{eff}}$ .

From the QSO catalog, lines-of-sight are drawn in the different boxes, and the fraction of transmitted flux, otherwise called **transmission**, is calculated using Eqn. 3.9.

### Underlying matter field

It is sometimes useful to compute the matter field associated to the Ly $\alpha$  forest lines-of-sight. Since, in these mocks, the matter distribution is a log-normal field, its associated contrast reads [6, 13]

$$\delta_m = \exp\left(\delta_L - \frac{\sigma_L^2}{2}\right) - 1, \quad (3.11)$$

where  $\sigma_L = \sigma(\delta_L)$  is the computed standard deviation of  $\delta_L$ . For this field, the minimal value is exactly  $-1$ . Since  $\delta_L$  follows a gaussian distribution of mean zero, the mean value of  $\delta_m$  is also equal to zero:

$$\langle \delta_m + 1 \rangle = \frac{1}{\sqrt{2\pi}\sigma_L} \int_{\mathbb{R}} \exp(\delta_L) \exp\left(-\frac{\sigma_L^2}{2}\right) \exp\left(-\frac{\delta_L^2}{2\sigma_L^2}\right) d\delta_L = 1. \quad (3.12)$$

Like  $\delta_L$ , the  $\delta_m$  field represents the matter density field at  $z = 0$ . To obtain it at any redshift, I multiply it by the same growth function  $G(z)$  as used in the FGPA approximation (Eqn. 3.9):

$$\delta_{\text{m}} = \exp \left( G(z) \times \delta_{\text{L}} - \frac{\tilde{\sigma}_{\text{L}}^2}{2} \right) - 1. \quad (3.13)$$

Here, the new standard deviation  $\tilde{\sigma}_{\text{L}}$  is defined to maintain the matter density field properties such that

$$\tilde{\sigma}_{\text{L}} = \sigma(G(z) \times \delta_{\text{L}}). \quad (3.14)$$

### 3.1.2 DESI-Lite mocks

Requirements for synthetic data are different to study the one-dimensional Ly $\alpha$  forest power spectrum  $P_{\text{1D},\alpha}$  detailed in Chap. 5. Given they were initially designed for BAO studies such as [9], the `SaclayMocks` did not model redshifts larger than  $z = 3.5$ . However, for the studies of Chap. 5, larger redshifts are important to improve the cosmological constraints. Furthermore, the true one-dimensional power spectrum of the `SaclayMocks` lines-of-sight is not defined analytically due to the different transformations performed on the GRFs. Thus, obtaining the true  $P_{\text{1D},\alpha}$  on these mocks is not possible without introducing an unwanted estimation step. Most importantly, while the production of 3D `SaclayMocks` takes a lot of CPU, it is convenient to have a cheaper way, in CPU time, to generate one-dimensional mocks.

The particular needs of the  $P_{\text{1D},\alpha}$  measurement led to the creation of dedicated mocks called `DESI-Lite` [14]. The full three-dimensional clustering is not needed, the lines-of-sight can be considered independent from each other which simplifies the mock creation.

Similarly to `SaclayMocks`, `DESI-Lite` are log-normal mocks based on a GRF. For each line-of-sight, a one-dimensional GRF  $\delta_{\text{1D,GRF}}(v)$  is generated in velocity space with a regular space gridding  $d_v$ . This GRF is used to compute the following baryon contrast

$$\delta_{\text{b}} = \mathcal{F}^{-1} \left[ \sqrt{\frac{P(k)}{d_v}} \times \mathcal{F}[\delta_{\text{1D,GRF}}(r)] \right], \quad (3.15)$$

where the input redshift-independent power spectrum is given by

$$P(k) = \frac{(k/k_0)^{n-\alpha \ln(k/k_0)}}{1 + (k/k_1)^\gamma}. \quad (3.16)$$

The parameters used are  $k_0 = 0.001 \text{ s}\cdot\text{km}^{-1}$ ,  $k_1 = 0.04 \text{ s}\cdot\text{km}^{-1}$ ,  $n = 0.5$ ,  $\alpha = 0.26$ , and  $\gamma = 1.8$ . The shape of the input power spectrum and the values of these parameters are obtained from fitting measured  $P_{\text{1D},\alpha}$  in [15]. The fitting function takes into account the redshift dependency. This dependency is added when calculating the fraction of transmitted flux with the log-normal approximation:

$$F = \exp \left[ -0.55 \left( \frac{1+z}{1+z_0} \right)^{5.1} \times \exp \left( 2a(z)\delta_{\text{b}} - a^2(z)\sigma_{\delta_{\text{b}}}^2 \right) \right], \quad (3.17)$$

where  $z_0 = 3.0$  is the reference redshift for the fit,  $\sigma_{\delta_{\text{b}}}$  is the measured standard deviation of the  $\delta_{\text{b}}$  field, and  $a(z)$  is defined by

$$a^2(z) = 58.6 \left( \frac{1+z}{1+z_0} \right)^{-2.82}. \quad (3.18)$$

Consequently, the average fraction of transmitted flux  $\bar{F}$  can be calculated analytically. Furthermore, using the definition of Ly $\alpha$  contrast in Eqn. 2.22, the input one-dimensional Ly $\alpha$  forest power spectrum is also known. The details of this calculation and the tests performed on DESI-Lite mocks are given in [14]. After generating the lines-of-sight, the velocity grid is rebinned to  $d_v = 30 \text{ km}\cdot\text{s}^{-1}$ . The minimal and maximal wavelengths are respectively 3600 – 9800 Å and the redshift distribution follows the expected DESI one [16, 17].

In conclusion, the DESI-Lite mocks provide quickly transmission files whose true one-dimensional power spectrum is known semi-analytically.

### 3.1.3 Modeling systematics in mocks

The transmission files generated with `SaclayMocks` or `DESI-Lite` have a common format. The `quickquasars` [18] software of the `desisim`<sup>3</sup> package, transforms transmissions into spectra with observational and astrophysical systematics. The complete pipeline to generate mocks is schematized in Fig. 3.1.

The `quickquasars` software first multiplies the transmissions by a continuum to produce a realistic quasar spectrum. This continuum is generated with the `simqso`<sup>4</sup> package. The underlying model implemented in `simqso` is described in [19]: each quasar continuum is constructed by adding a set of emission lines on top of broken power laws.

The code is used to add HCD systems to the spectra. To obtain a reasonable HCD clustering, it is necessary to generate them from the matter field associated with the mocks. For the `SaclayMocks`, the HCD objects are added where the large-scale matter field  $\delta_L$  is higher than a given threshold. The probability of generating HCD objects in these regions follows a Poisson law. The distribution of HCD column density  $N_{\text{HI}}$  follows eBOSS measurements [20]. For the `DESI-Lite` mocks, HCD clustering is not required. HCD systems are placed randomly on the lines-of-sight following the  $N_{\text{HI}}$  distribution. BAL can also be added to the QSO spectra using `quickquasars`. Finally, the software models the absorption due to metals present in the IGM for a given list or all metals.

The instrumental systematics are also modeled. The noise is determined from a fictitious exposure time and a maximum observed magnitude. This modeling is calibrated to reproduce the pipeline noises of SDSS or DESI, introduced in Chap. 4. To mimic eBOSS spectral resolution, a Gaussian smoothing and a resampling into a wavelength grid are applied. For DESI, the resolution is modeled by a constant resolution matrix.

The final software outputs are spectra and an associated quasar catalog. Depending on the Ly $\alpha$  analyses one wants to perform, the list of systematics taken into account in the mock spectra will be different.

3. <https://github.com/desihub/desisim>

4. <https://github.com/desihub/simqso>

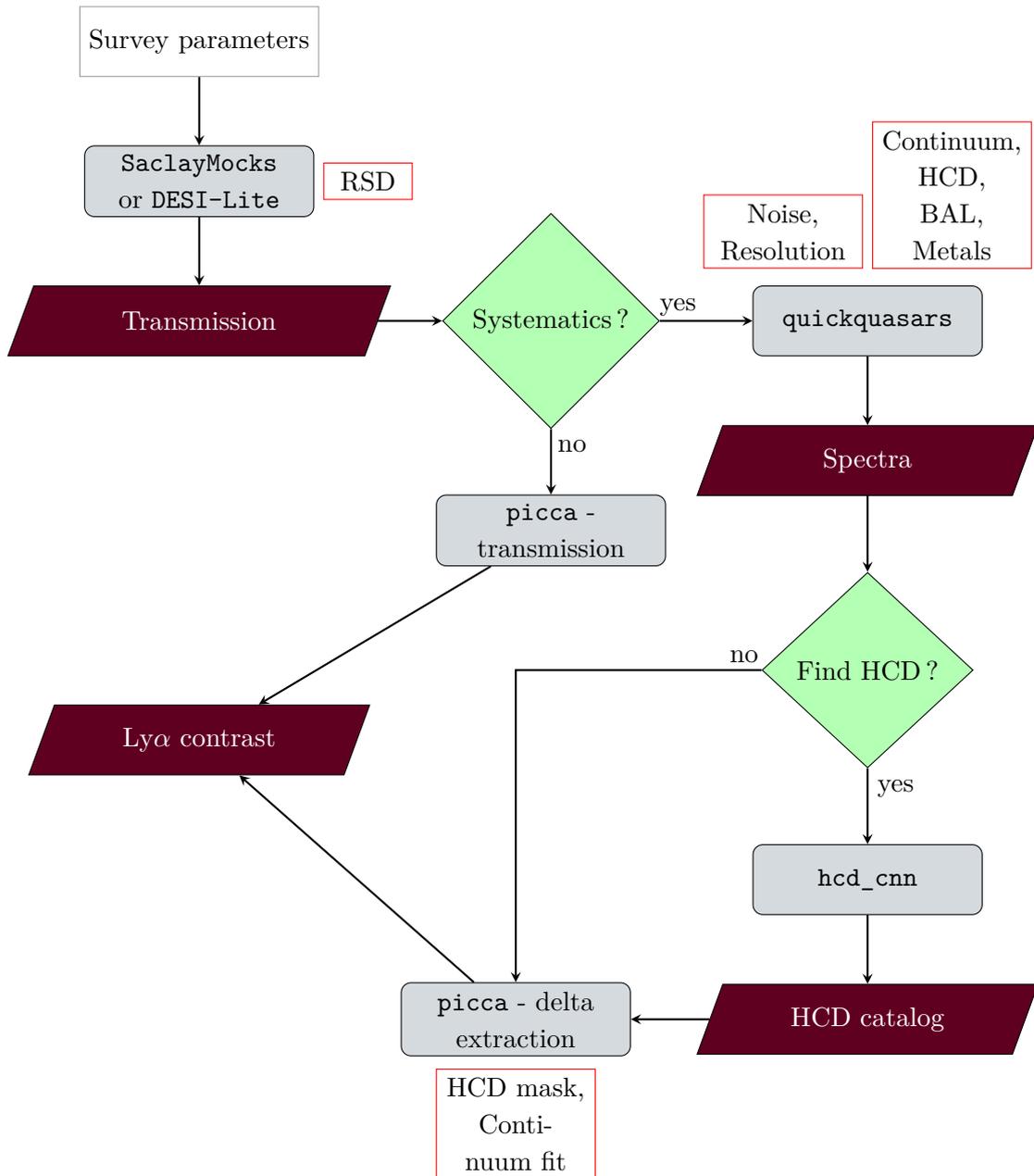


FIGURE 3.1 – Schematic representation of the mock creation and processing used to model the Ly $\alpha$  systematics. The creation of files containing the transmissions can be done by either `SaclayMocks` or `DESI-Lite`. Once the spectra are created, the pipeline which computes Ly $\alpha$  contrasts is the same as for real data.

## 3.2 Cosmological simulations

Future cosmological surveys will provide high-precision data. In this context, the theoretical predictions must achieve an equivalent level of accuracy. Approximated mocks are generally not sufficient to fully characterize the impact of cosmological parameters on the matter field. Cosmological simulations are then required.

With the exponential increase in computing resources, the field of computational cosmology has undergone significant growth. One of the first simulations [21] now has more citations than the number of dark matter particles it included. Current large simulations such as *PkDgrav3* [22], *HorizonAGN* [23], *AbacusSummit* [24], or *OuterRim* [25] use an extremely large number of particles or resolution elements ( $> 10^{10}$ ). Many cosmological or astrophysical phenomena can be modeled in these numerical laboratories. The realization of these simulations requires extensive use of large supercomputers and constitutes an important challenge in memory usage and computation time. In order to optimize them, specific implementations on graphical processing units (GPUs) are now used. In this context, I have contributed to testing the **Nyx** software on a recent GPU partition.

In Sec. 3.2.1, a general overview of cosmological simulations is presented. I contributed to design cosmological simulations with the *Nyx* software to model both one-dimensional (Sec. 3.2.3) and three-dimensional Ly $\alpha$  forest power spectra (Sec. 3.2.4). These two grids of simulations use a common pipeline described in Sec. 3.2.2.

### 3.2.1 Overview

Cosmological simulations are used to simulate dark matter, baryons, or neutrinos by solving equations presented in Sec. 1.3.3. They are expressed in comoving coordinates and model the evolution of a fixed comoving volume as a function of redshift. Here, we consider simulations on a cubic volume with periodic boundary conditions (to suppress edge effects).

#### Solving gravity

As show in Sec. 1.3.3.1, the Vlasov equation in the expanding universe 1.116 associated with the Poisson equation 1.115 forms a system called **Vlasov-Poisson**. This system governs the evolution of gravity in the Newtonian framework. Cosmological simulations aim to solve this equation using different numerical methods. They can be categorized by the discretization of the phase space in which the Vlasov-Poisson system is expressed. A graphical representation of the different types of discretizations is shown in Fig. 3.2.

For our purpose, the commonly used method is the **N-body technique** in the Newtonian approximation. With this method, the represented cosmological fluid is sub-sampled into point-like particles of fixed mass. By imposing a gravitational integration, it is possible to evolve the position and velocity of particles until a given redshift. The properties of the cosmological fields, such as density, can be recovered by interpolating the particles on a grid.

Two other classes of numerical methods, called geodesic deviation equation (GDE) and sheet tessellation, track the field properties along with the particles. The GDE approach (see e.g. [26, 27]) associates a tensor for each particle and follows the phase space deformations. The second method consists in tessellating the phase space from the particles, which are considered as vertices [28]. These two methods give a better representation of the phase space deformation but are very time-consuming.

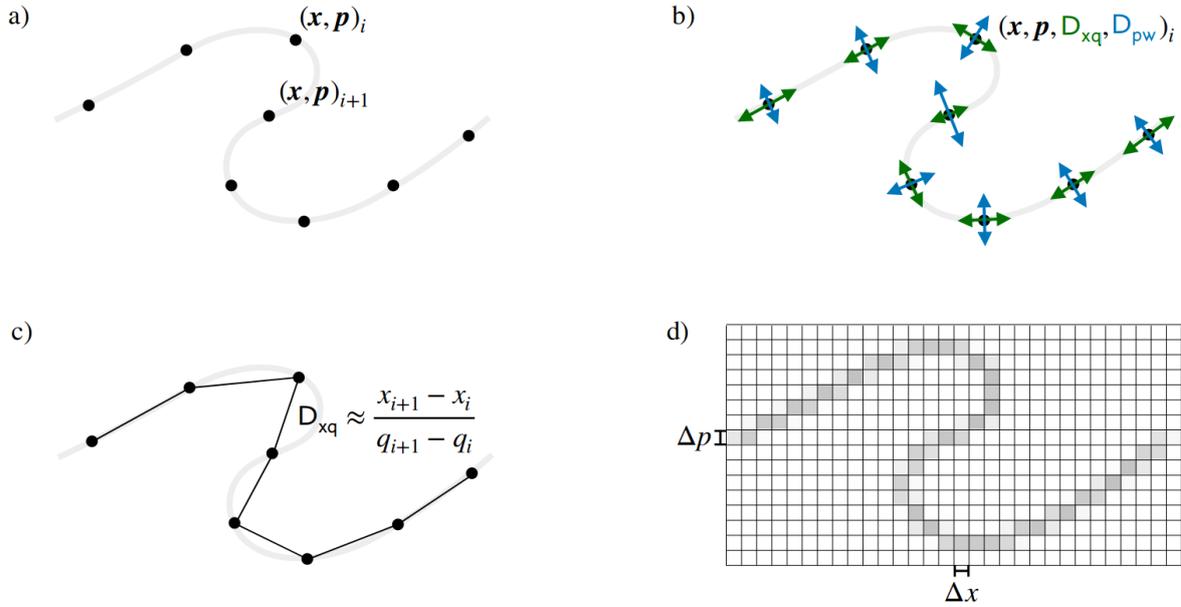


FIGURE 3.2 – Graphical representation of the different techniques used to discretize the Vlasov-Poisson system, taken from [1]. (a) The  $N$ -body method samples the cosmological fields into particles. (b) The geodesic deviation equation tracks manifold deformation by associating a tensor to each particle. (c) The sheet tessellation method approximates manifold deformation by interpolating particles. (d) The finite volume discretization discretizes phase space over a regular grid.

The last type of method called Vlasov-Poisson simulations [29, 30] consists in solving Vlasov-Poisson system of equation by directly integrating the distribution function  $f(\vec{x}, \vec{p}, t)$  in 6-dimensional space. For this, the position and velocity fields are discretized over a regular grid, as shown in Fig. 3.2. These methods are extensively memory-consuming and better adapted to systems with a small dynamic range.

### **$N$ -body simulations: dark matter**

In  $N$ -body simulations, the matter is represented by sampled points of the phase space distribution function called particles. This approach is fully equivalent to simulating the motion of  $N$  particles. The larger the sampling, i.e., the number of particles, the better the simulation resolution. The  $N$ -body approach is appealing by its simplicity but has a major disadvantage. The equation to solve 1.116 is collisionless. When two particles in a  $N$ -body simulation collide, this property is not respected and leads to artificial two-body relaxation. This issue is alleviated by imposing a gravitational softening that smooths the density field at small scales. The time is also discretized during the simulation, and the particle trajectory calculation is performed at each time step. When the particles are very close, and the interaction forces become large, a variable time stepping must be used to maintain the precision of particle's motion.

There are several sub-classes of  $N$ -body simulations depending on the numerical scheme used to compute the gravitational force. The **particle-particle (PP)** method is the simplest but also the most expensive in terms of computational time (complexity in  $\mathcal{O}(N^2)$ ). In this case, the gravitational potential is calculated by direct summation of all particle contributions.

The second subclass called **particle-mesh (PM)** consists in solving the Poisson equation on a regular grid called mesh. The matter density is calculated on the mesh at each step by

interpolating the particle positions. On the mesh, the computation of the Poisson equation 1.115 in Fourier space is accelerated using a fast Fourier transform (FFT). It reduces the complexity of PM implementation to  $\mathcal{O}(N \log(N))$ . The gravitational potential is used to derive the forces imposed on the particles. Several methods are possible to interpolate the density field. The nearest-grid-point, cloud-in-cell, and triangular-shaped-cell methods respectively use one, two, or three points along each direction to interpolate the density.

The **tree-code** method organizes the particles of the simulation along a tree. The most expensive term to compute in  $N$ -body simulations is the large-scale gravitational interaction called far-field. In the tree-code implementation, the small-scale gravitational forces are computed by taking into account only the nearest neighbors of each particle. The far-field forces are calculated using the tree computed from particle separations. In the tree's larger branches, the total mass and the center of mass give a good approximation of this far-field force. By summing the far-field contribution to the nearest neighbor forces, the complexity of this method drops to  $\mathcal{O}(N \log(N))$ . The differences in implementing this approach are mainly related to the tree creation technique. It is possible to distinguish between bottom-up and top-down schemes, for which the particles of the simulation are respectively the leaves or roots. The computation of the far-field forces can be done by a hierarchical multipole expansion using a Fast-Multipole-Method scheme which improves the accuracy of the tree-code method.

The codes we work with (**Gadget**, **Nyx**) use more complex hybrid methods like the **particle-particle/particle-mesh (P<sup>3</sup>M)** which couples a PP calculation for small-scales forces and PM for far-field. The **tree-code particle mesh (TreePM)** method uses a tree-code method in the very dense regions and a PM in the underdense regions.

### Hydrodynamical simulations: baryons

Many cosmological simulations do not require to model baryons. Indeed, the dark matter dominates the matter Universe's composition and thus its large-scale structuring. In particular, for large galaxy surveys,  $N$ -body simulations are used with calibrated models to draw galaxies from the dark matter field (halo and occupation distribution models). The pressure-less behavior of dark matter greatly simplifies the implementation of these simulations. On the other hand, for numerical modeling of the Ly $\alpha$  forest, the baryon pressure and temperature in the smooth, continuous IGM must be considered. It is then necessary to add equations for the baryon fluid, i.e., the system formed by the continuity, Euler, and Poisson equations 1.120. Two main numerical methods are used to solve this problem. They are based on the Eulerian and Lagrangian approaches introduced in Sec. 1.3.3.

The **Eulerian method** treats the gas properties on a grid. To have a finer resolution in the densest regions, an adaptive mesh refining (AMR) scheme can be used. The equation set 1.120 is usually solved through a finite difference scheme. Another type of implementation called **Lagrangian method** is analog to the  $N$ -body approach for modeling baryons. In this approach, the gas is modeled by particles whose trajectories are followed. Unlike  $N$ -body simulations, the thermodynamical properties of the gas are associated with each particle. This method is also called smooth-particle hydrodynamics (SPH). Local gas properties, such as temperature and density, are computed by convolving the particle properties with a smoothing function, for example, Gaussian or cubic kernels.

Both methods have advantages and drawbacks, and their use depends on the application. The SPH method is adapted to maximize the resolution in overdense areas. However, it is not very efficient for modeling shocks or underdense regions. In practice, interpolating the baryonic

properties along Ly $\alpha$  lines-of-sight from SPH snapshots is also not straightforward. The Eulerian method is more adapted to GPU programming. The post-processing of these simulations is easier. However, it is sensitive to large velocities, and the overdense regions can be under-resolved. For the Ly $\alpha$  forest modeling, it is appropriate to have a uniform resolution in both under- and overdense areas so that an AMR method is not needed. The **Gadget** code uses SPH, while **Nyx** uses an Eulerian grid. Our group previously used **Gadget** simulations for BOSS analysis, but in this thesis, I mostly used **Nyx** which provides significant computational gains.

In all hydrodynamical simulations, it is essential to model the heating and cooling processes of the IGM. As indicated in Sec. 2.1.2, IGM thermodynamic equilibrium is controlled by the competition between ionization and recombination phenomena. Those play an essential role in determining the IGM temperature, which significantly influences the Ly $\alpha$  forest. These effects are modeled using tabulated cooling functions, which take into account a model for the Universe thermal history (UV background, reionization...).

Simulating the Ly $\alpha$  forest does not require detailed knowledge of the overdense regions of the cosmic web. In particular, it is not necessary to model the formation of galaxies as the flux from a quasar would be fully absorbed. To simplify the simulations, a density criterion is generally applied: the gas condensation processes are not modeled in these overdense regions.

Several additional sub-grid baryonic effects can be added to simulations. In particular, it is possible to include star formation processes, supernovae feedback, or AGN feedback. This comes at the price of a much higher CPU cost, since in this case galaxy formation must be simulated.

### Simulating neutrinos

The Ly $\alpha$  forest and, in particular, its one-dimensional power spectrum (see Chap. 5) is sensitive to the sum of the neutrino masses. Indeed, as shown in Sec. 1.3.4.1, massive neutrinos tend to reduce structure formation at scales measured by this observable. Numerical modeling of neutrinos is challenging because of their high velocities.

The Vlasov-Poisson approach solves this problem as high velocities are naturally taken into account in this scheme [31, 32]. In  $N$ -body simulations, several approaches are used to tackle with this issue. One method uses the **quasi-linear evolution** of neutrinos and models them only with a Boltzmann solver [33, 34, 35]. At each step of the  $N$ -body simulation of dark matter, a correction is applied to the Boltzmann hierarchy to include the impact of neutrinos. This method may be coupled to switching to  $N$ -body neutrino modeling when their clustering becomes important. The **Fourier-space methods** [36, 37] add the impact of neutrinos on the mesh of PM simulations. The gravitational force of neutrinos is calculated in Fourier space and directly added to the gravitational potential. A method called **linear-response approximation** (LRA) [38] treats in a semi-linear way the neutrino impact. Assuming that neutrinos and dark matter overdensities are in phase, the neutrino power spectrum is computed at each step and used to correct the dark matter particles displacements. Finally, the neutrinos can be directly modeled by particles. In this case, special treatments of **neutrino initial conditions** [39, 40] are needed to reduce the numerical impact of high-velocity neutrinos. Hybrid methods like [41] use LRA for the first step of the simulation to avoid the previous numerical effects and model neutrinos as  $N$ -body particles at low redshift when they start clustering.

In another approach, adapted to the Ly $\alpha$  forest at least in first approximation, one can mimic the impact of neutrino masses in a simulation without neutrinos: cosmological parameters are changed to reproduce the effect of  $\sum_{\nu} m_{\nu}$  on the amplitude and slope of the matter power

spectrum in the  $(k, z)$  range that is relevant for Ly $\alpha$  [42, 43, 44]. This simplified approach works because the non-linear effects intrinsic to  $\sum_{\nu} m_{\nu}$  are small.

### 3.2.2 Simulation pipeline

The simulation pipeline used for this chapter is described here and schematically represented in Fig. 3.3.

#### Initial conditions

The initial power spectrum of a simulation is computed with the CLASS<sup>5</sup> [45, 46] Boltzmann solver. An example of matter power spectrum obtained with CLASS is given by Fig. 1.12. The power spectrum depends on the cosmological parameters detailed in Sec. 1.2. These parameters are imposed again in the simulation. A flat  $\Lambda$ CDM model is used for all the simulations of this chapter.

Once the matter power spectrum is computed, it is necessary to generate **initial conditions (IC)** at a redshift  $z_{\text{start}}$ , in terms of particles or fields depending on the simulation's type. This is done by solving the Lagrangian equations detailed in Sec. 1.3.3.2 with the nLPT approach. The redshift  $z_{\text{start}}$  is chosen high enough so that the IC generation method does not introduce errors. This redshift is generally comprised between  $z = 50$  and 150. The IC software defines the simulation geometry in terms of comoving box length. The number of particles or hydrodynamical voxels is also parameterized during the IC step and defines the simulation resolution. In this chapter, the IC software 2LPTic<sup>6</sup> [47] and monofonic<sup>7</sup> [48, 49] are used. They respectively implement the nLPT approximation at second and third order. For SPH hydrodynamical simulation, the baryons and dark matter particles can be initialized in the same way. For the Eulerian method used in this chapter, initial conditions for baryons can be generated directly or by interpolating particles on a grid. With the version of 2LPTic we used, we generated dark matter particles from the total matter transfer function. The initial baryon grid was then initiated by subtracting a fraction of the particle masses and depositing it on a mesh using a cloud-in-cell (CIC) scheme. Therefore, the same transfer function was used for baryons and dark matter when running 2LPTic. Using this approximation generates a bias in the predicted Ly $\alpha$  power spectrum, especially at high redshift [50]. Therefore, we also run ICs with monofonic. In that case, SPH and dark matter particles are generated with their respective transfer functions. The SPH particles are then deposited into the grid with the same CIC method.

#### Simulation code

In this chapter, simulations are performed with the hydrodynamical grid code Nyx<sup>8</sup> [51, 52]. This software models baryons as an ideal gas on a Eulerian grid. The gas dynamics is solved using a second-order accurate finite volume methodology. Nyx has AMR capabilities, but this is not used here as the Ly $\alpha$  forest needs to be simulated with a uniform spatial resolution. The mesh structure used to evolve fluid quantities is also used for the PM  $N$ -body scheme. Besides solving for gravity and the Euler equations, Nyx includes the physical processes essential to modeling the Ly $\alpha$  forest. The chemistry of the gas is considered as having a primordial fixed composition with

---

5. <http://class-code.net>

6. <https://cosmo.nyu.edu/roman/2LPT>

7. <https://bitbucket.org/ohahn/monofonic>

8. <https://amrex-astro.github.io/Nyx>

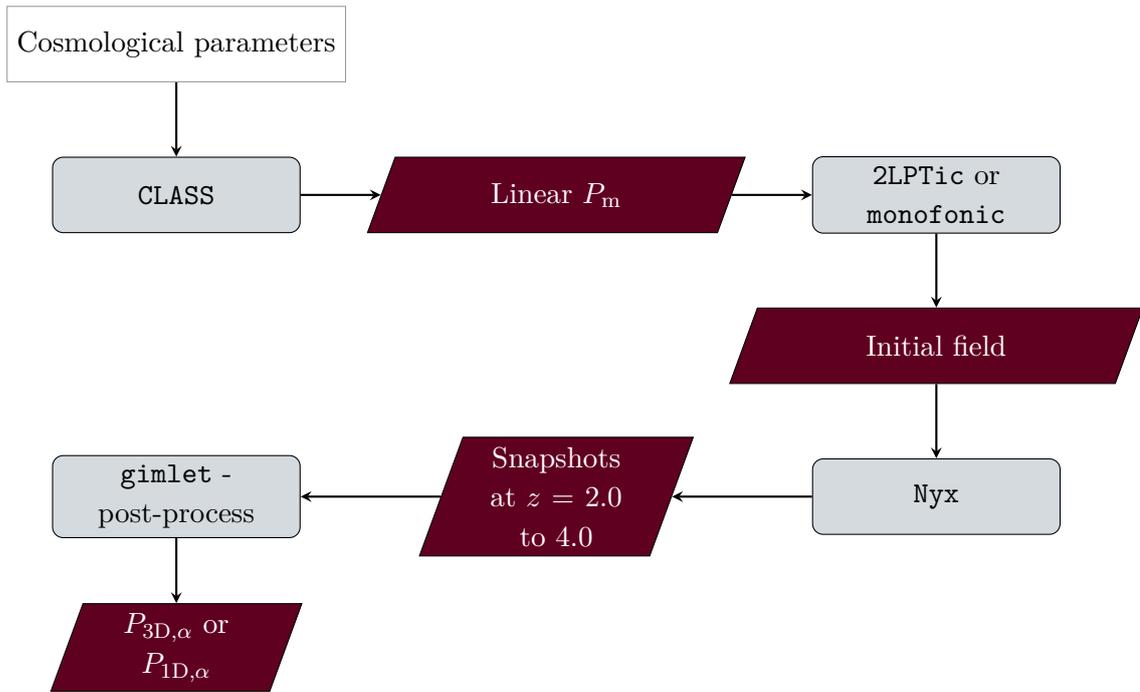


FIGURE 3.3 – Diagram of the pipeline used for the two simulation sets presented in Sec. 3.2.3 and 3.2.4.

hydrogen and helium mass abundance. In addition, *Nyx* models the inverse Compton process, and keeps track of the net loss of thermal energy induced by atomic collisional processes. The recombination, ionization, and cooling rates are defined according to [52]. All cells are assumed to be optically thin to ionizing radiation, and radiative feedback is accounted for via a spatially uniform but time-varying UV background whose redshift evolution is given by hand. In all the simulations, neutrinos are not modelled.

### Sub-grid physics

*Nyx* is a relatively recent software and turns out to be adapted to model the physics of the Ly $\alpha$  forest quickly. However, the sub-grid physics implemented is insufficient to simulate the IGM state fully. Indeed, *Nyx* does not explicitly model galaxy formation and does not follow the radiative transport through the box. This impacts the Ly $\alpha$  forest in several ways:

- **Galaxy formation feedback:** *Nyx* does not account for thermal feedback of AGN or supernovae in galaxies. Recent works on simulations like Horizon-AGN [53] have shown that AGN feedback consequently modifies the matter power spectrum [54] and the one-dimensional Ly $\alpha$  forest power spectrum [55], and this over an extended dynamical range. However, this type of simulation requires significant computing resources and often requires AMR to reach smaller scales.
- **Uniform radiative background:** For larger scales, the inhomogeneity of the radiative transport is not accounted for but modeled by a uniform UV background. This simplification can impact the Ly $\alpha$  forest for all redshifts. Studies have already been done to characterize the impact of this effect on the correlation functions in [56].
- **Reionization:** Finally, the history of reionization is modeled uniformly and with a fixed model. Changing this history mainly impacts the Ly $\alpha$  forest for the highest redshifts. Some studies focus more specifically on the modeling of reionization [57], such as Cosmic Dawn

II [58, 59] and Cosmic Dawn III (in production), which explicitly simulate the radiative feedback for high redshifts. However, simulations like Cosmic Dawn III ( $8192^3$  dark matter particles,  $8192^3$  hydrodynamical elements, and radiation cells) require running one of the largest supercomputers in the world for several tens of days both with GPU (radiation) and CPU (baryons and dark matter) to reach redshifts  $z \sim 4$ . Some recent works focus on characterizing this effect for Ly $\alpha$  forest [60].

It would be desirable to use simulations including sub-grid physics explained above to model the Ly $\alpha$  forest fully, but this is hardly accessible for the actual high performance computing capacities: for simulations in Sec. 3.2.3, we wish to run several realizations to model the impact of cosmological parameters.

Consequently, we choose to run our simulations without explicitly simulating these effects, but we take them into account at the fitting stage. For example, the AGN feedback impact will be marginalised for the one-dimensional power spectrum measurement (see Chap. 5) thanks to the separate simulations done in [55]. It is clear that other simulations still must be performed to calculate the impact of all sub-grid effects on the measured observables.

### Post-processing

Once the baryons and the dark matter distributions are computed at the redshift of interest, simulation outputs (snapshots) are post-processed using a modified version of the `gimlet`<sup>9</sup> suite [61]. This software package allows for efficient, MPI-parallel computation of typical statistical properties in a Nyx simulation based on the hydrodynamical grid and CIC depositions of the dark matter particles. In particular, it computes the one-dimensional and three-dimensional Ly $\alpha$  forest power spectra. In this case, the optical depth is calculated according to Eqn. 2.15. The absorption field is Fourier transformed in one or three dimensions depending on the observable needed. The one-dimensional power spectrum is obtained by averaging the power estimated from skewers drawn along the three simulation directions ( $X, Y, Z$ ). For  $P_{1D,\alpha}$  on small scales, this is equivalent to using three independent simulation boxes and reduces sample variance accordingly.

## 3.2.3 Simulating the one-dimensional Lyman- $\alpha$ forest power spectrum

The interpretation of the measured one-dimensional power spectrum (Chap. 5) requires modeling the IGM down to  $\sim h^{-1}$ .Mpc scales. This can only be done with hydrodynamical simulations. Furthermore, fitting  $P_{1D,\alpha}$  requires to predict this observable for different combinations of cosmological parameters. Given the required accuracy and typical supercomputer allocations, only a limited number of simulations can be performed. Consequently, some interpolation scheme is needed to compute  $P_{1D,\alpha}$  from those simulations for arbitrary cosmological parameters.

This section will briefly summarize the construction and performance of such a simulation grid dedicated to  $P_{1D,\alpha}$ . All the simulation run in this section are initialized with the 2LPTic software at  $z_{\text{start}} = 99$ . More details on the  $P_{1D,\alpha}$  grid and associated tests are given in [4]. I contributed to this work by running some MP-Gadget<sup>10</sup> simulations for matter power spectrum comparison tests, and by helping to run the main Nyx grid.

9. <https://github.com/bcfriesen/gimlet>

10. <https://github.com/MP-Gadget/MP-Gadget>

### Emulator

Previously, for cosmological parameter estimation from the BOSS  $P_{1D,\alpha}$ , the interpolation scheme was done from a fiducial simulation, along with additional simulations that each modified one or two of the input parameters. These simulations were used to build a first-order [62] or second-order Taylor expansion including cross-term contributions [63]. The Taylor expansion approach ensures a good accuracy near the fiducial model and along the expansion axes but does not cover with homogeneous precision the parameter space away from this fiducial model. A **Gaussian process emulator** is alternatively used to overcome this issue. Such an interpolation was recently applied to multiple Ly $\alpha$  forest analyses [43, 64, 65, 66].

For our emulator, a grid of simulations is generated with parameters  $\Theta_{\text{sim}}$  arranged in a space-filling Latin hypercube design (LHD). A Gaussian process is then trained to emulate each of the  $P_{1D,\alpha}(z, k)$  modes for new sets of parameters  $\Theta_{\text{emu}}$ . The main advantage of this approach is a more uniform interpolation scheme that does not rely on a single fiducial model. Furthermore, it is possible to get an estimate of the interpolation uncertainty on-the-fly without running additional simulations. Also, one can increase the accuracy in a specific region of parameter space by running additional simulations.

For the cosmological parameter space, instead of using the base  $\Lambda$ CDM parameters, we adopt a derived parameterization best adapted to the  $P_{1D,\alpha}$  measurement. In particular, the  $(A_s, n_s)$  are replaced by

$$\begin{aligned} A_{\text{Ly}\alpha} &= P_m(k = 1 \text{ Mpc}^{-1}, z = 3), \\ n_{\text{Ly}\alpha} &= \frac{\partial \ln P_m}{\partial \ln k}(k = 1 \text{ Mpc}^{-1}, z = 3), \end{aligned} \quad (3.19)$$

where  $P_m$  is the linear matter power spectrum. Using CLASS [45, 46], one can compute  $(A_{\text{Ly}\alpha}, n_{\text{Ly}\alpha})$  from  $(A_s, n_s)$  and reciprocally.

For this emulator, the UV heating rates of the simulation  $\epsilon$  are rescaled, which is the standard approach when performing Ly $\alpha$  forest modeling (see e.g. [67, 68]). This procedure provides different thermal histories and thus different amounts of pressure smoothing. The heating rates implemented in the simulation is

$$\epsilon = A_{\text{UVB}} \left( \frac{\rho_b}{\bar{\rho}_b} \right)^{B_{\text{UVB}}} \epsilon_{\text{fid}}, \quad (3.20)$$

with  $\epsilon_{\text{fid}}$  being the fiducial heating rates and  $(A_{\text{UVB}}, B_{\text{UVB}})$  are simulation parameters.

The final set of simulation parameters is  $\Theta_{\text{sim}} = \{A_{\text{Ly}\alpha}, n_{\text{Ly}\alpha}, A_{\text{UVB}}, \omega_{\text{m},0}, H_0\}$ , implying that  $\omega_b$  and  $B_{\text{UVB}}$  are the same for all simulations. 14 simulations were run, with LHD points represented in Fig. 3.4. The cosmological parameters span a  $\pm 5 \sigma$  range around the Planck best fit values [69].

A two-fold approach is used to vary the IGM thermal state. First, to generate different thermal histories, we rescaled UV heating rates by the  $A_{\text{UVB}}$  factor. This rescaling effectively changes the temperature at mean density  $T_0(z)$ , defined by Eqn. 2.3. However, the density dependence of each simulation characterized by  $\gamma$  is unchanged. A post-processing procedure at every output redshift is applied to modify the instantaneous temperature-density relation. We rescale temperatures of all hydro cells by a factor  $T_0/T_{0,\text{sim}} (\rho_b/\bar{\rho}_b)^{\gamma-\gamma_{\text{sim}}}$ , such that a new instantaneous temperature-density relation  $T = T_0 (\rho_b/\bar{\rho}_b)^{\gamma-1}$  is generated while pressure smoothing is unchanged, allowing for independent modification of both effects. Finally, we generate different mean transmissions  $\bar{F}$  in the Ly $\alpha$  forest by rescaling the optical depth in post-processing. For

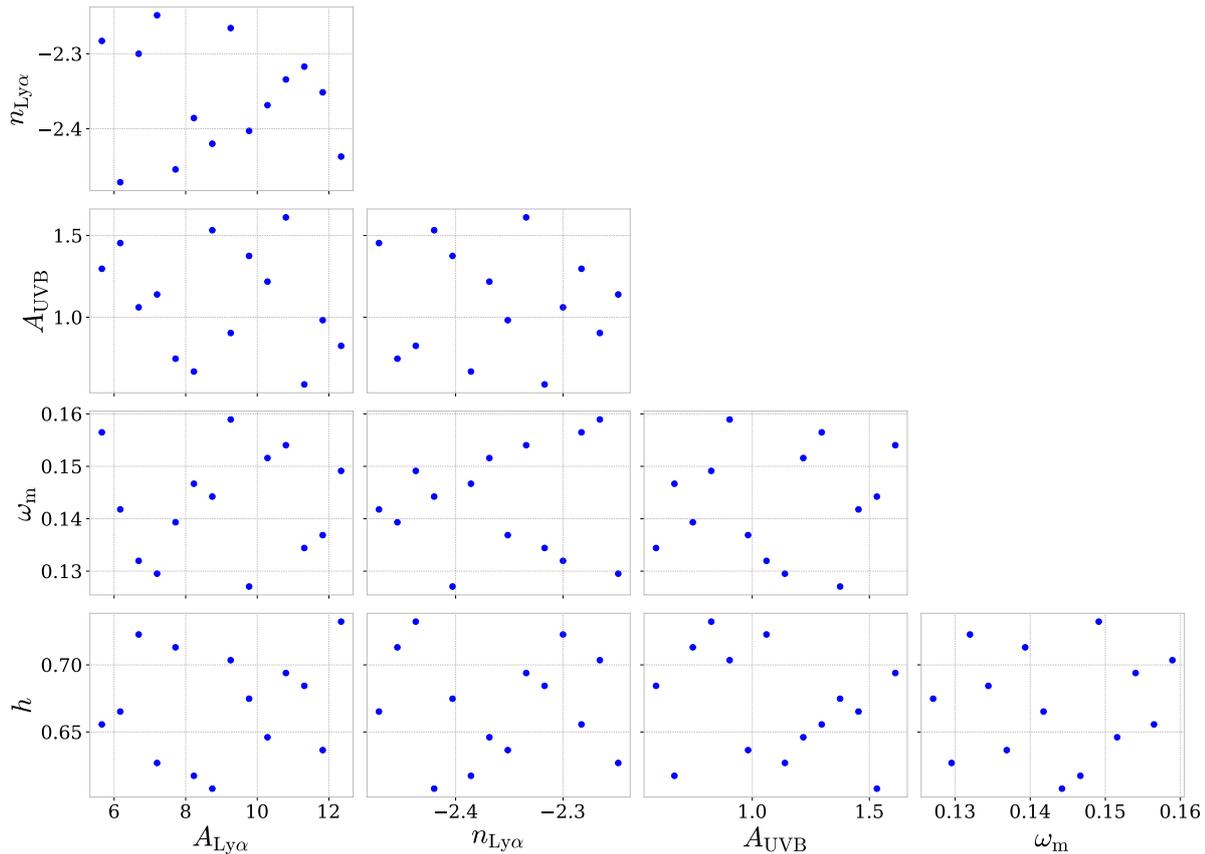


FIGURE 3.4 – The points generated by the Latin hypercube design  $\Theta_{\text{sim}}$  which correspond to the cosmological and UV background parameters of the fourteen simulations. The LHD maximizes the coverage of the parameter hyperspace [4].

each redshift, the values of  $\Theta_{\text{post}} = \{T_0, \gamma, \bar{F}\}$  were chosen from a second, denser LHD. To summarize, we generated one LHD covering the five simulation parameters and one LHD for each output redshift covering the three post-processing parameters.

The Gaussian process emulator is built individually for each redshift, which gives the best performance, using the combined set of  $\Theta_{\text{sim}}$  and  $\Theta_{\text{post}}$  as input parameters. This emulator is trained to compute  $P_{1D,\alpha}$  with associated uncertainties for every point in the parameter space covered by the two LHD. The Gaussian process was created using the python package *George* [70].

### Final Nyx grid

The expected DESI observations [17] indicate that it will be possible to achieve, at the survey's end, a measurement of  $P_{1D,\alpha}$  with a near- $\%$  precision. In order to interpret this measurement, it is necessary to reach a similar accuracy level on the estimation of  $P_{1D,\alpha}$  in the simulations. Independently of requirements concerning the physics implemented in simulations (e.g. AGN feedback, reionization model), this first implies constraints on the resolution and grid size of the simulations. In [4], convergence tests of Nyx regarding resolution and box size are shown.

Regarding resolution, the DESI range of interest of the Ly $\alpha$  forest is defined on observational requirements. For  $z \leq 3.2$ , which is the redshift range where the most data are anticipated,

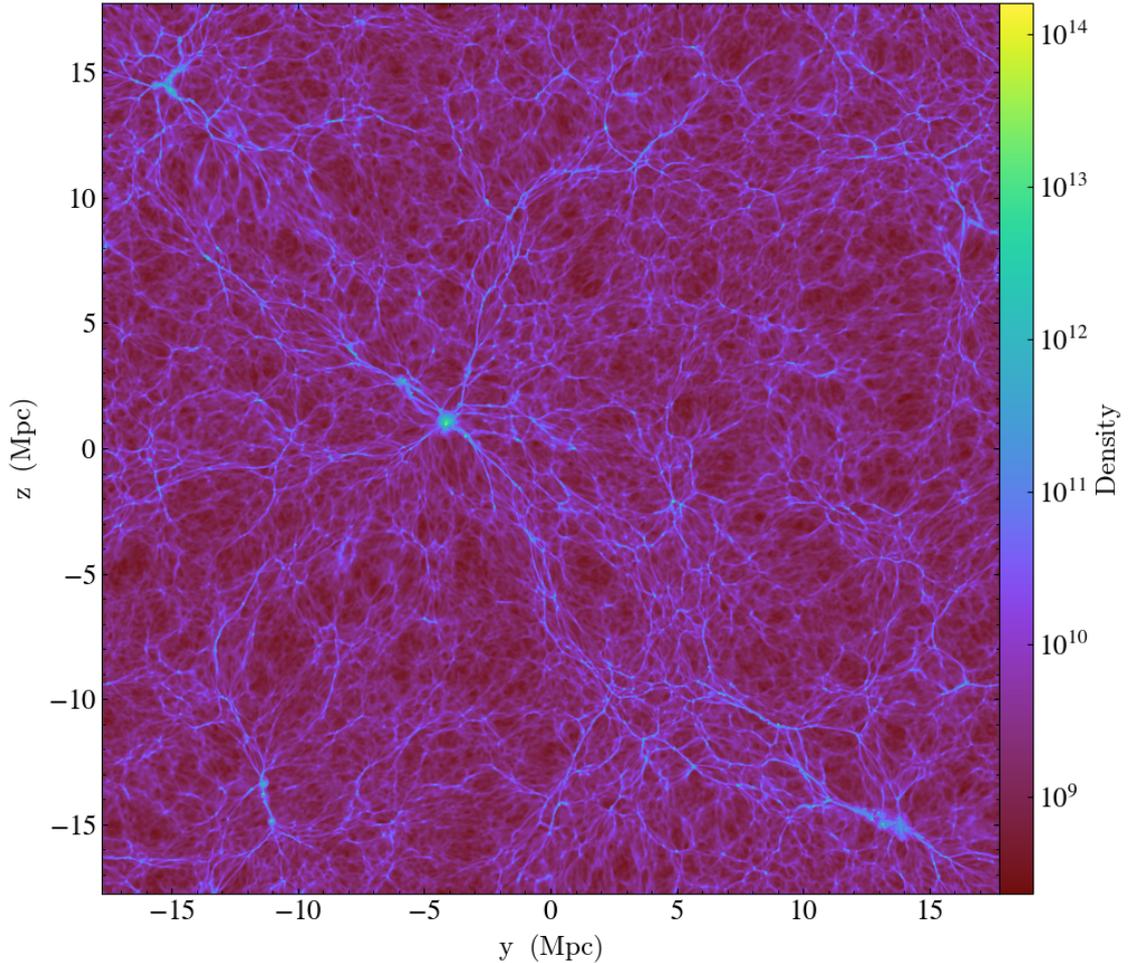


FIGURE 3.5 – Matter density slice plot of the fiducial simulation at  $z = 2.6$ . Planck best fit cosmological parameters [69] are used. The comoving size is 120 Mpc for each direction which corresponds to the physical size shown in the figure axis. Matter density is expressed in unit of  $M_{\odot} \cdot \text{Mpc}^{-3}$ .

resolution convergence must be  $\lesssim 1\%$  for the whole range of scales  $k < 0.1 \text{ s} \cdot \text{km}^{-1}$ . The latter wavenumber limit is fixed considering the DESI resolution. In [4], we show that simulations with a resolution lower than 30 kpc are needed to validate these requirements with Nyx.

The box size affects simulation in multiple ways. There is an intrinsic sample variance effect on the largest scales, which affects progressively smaller scales and increases in amplitude as box size decreases. In addition, lacking large-scale modes leads to underestimating bulk flow velocities of the gas, which in turn underestimate thermal broadening and RSD [52], which both would change the small scale power.

In order to achieve sub-percent level convergence over most of the scales and redshift relevant for the  $P_{1D,\alpha}$ , simulation boxes of 120 Mpc size are needed. In any case, boxes of at least this size are also required to access large-scale modes  $k \sim 0.001 \text{ s} \cdot \text{km}^{-1}$ .

In conclusion, for analyses of DESI data sets, at least 120 Mpc boxes with  $4096^3$  resolution elements need to be run, which is in line with previous works [52, 63]. Performing a single simulation with Nyx on modern CPU-based supercomputers takes approximately two million CPU hours. The fourteen points of the emulator grid described previously were calculated using

a PRACE allocation. The simulations are performed using the pipeline detailed in Sec. 3.2.2. As indicated in Sec. 3.2.2, the 2LPTic initialization used here implies that the baryons and the dark matter are not generated separately. An independent run with separated initial conditions was also performed to characterize the impact of ICs. A representation of the fiducial simulation density field is given in Fig. 3.5. The exploitation of this grid for cosmological constraints is currently in progress. It will be first used on eBOSS data to compare with previous results [71] and then employed for DESI  $P_{1D,\alpha}$  interpretation.

### 3.2.4 The three-dimensional power spectrum of the Lyman- $\alpha$ forest

The one-dimensional power spectrum is the only achieved small-scale, Fourier-space measurement of the Ly $\alpha$  forest with existing data. With the advent of DESI and WEAVE-QSO, the measurement of the three dimensional power-spectrum  $P_\alpha$  (noted  $P_{3D,\alpha}$  in this section) might be achievable. Indeed, due to the very anisotropic shape of the Ly $\alpha$  forest, this measurement is only possible for a large line-of-sight density that provides a similar spatial resolution in the transverse and radial directions. An example of a method to calculate  $P_{3D,\alpha}$  is given in [72].  $P_{3D,\alpha}$  was already calculated from hydrodynamical simulations in [73, 74, 75]. However, the simulation boxes used in these studies remain of limited sizes.

To compute  $P_{3D,\alpha}$  up to very large scales, I co-lead a collaborative project with researchers from DPhP, DAP, and LBNL to produce large-volume hydrodynamical simulations up to  $500 h^{-1}\cdot\text{Mpc}$ . Hydrodynamics is often neglected or under-resolved in existing simulations with large box sizes. These new simulations provide an ideal laboratory to test analytical models with sufficient accuracy while limiting cosmic variance effects. The main goals of this project were the following:

- **Nyx implementation:** Eulerian grid hydrodynamical codes such as Nyx are very well suited to be implemented on GPUs as their architecture facilitates the resolution of matrix problems. Our project contributes to the testing of a recent Nyx GPU implementation. In particular, the use of a large number of GPUs has highlighted problems that have been further corrected by the Nyx development team.
- **RSD modeling:** The RSD impact Ly $\alpha$  forest over a wide range of scales. However, this effect is strongly blended with nonlinear effects on scales smaller than about  $20 h^{-1}\cdot\text{Mpc}$ . Existing hydrodynamical simulations [74] are therefore severely limited by the box size to study RSD at the most relevant scales. A large volume of more than  $100 h^{-1}\cdot\text{Mpc}$  is required for the tidal fields to be representative. With large boxes, we aim at improving prediction for Ly $\alpha$  fluctuations in the linear regime, for which the RSD effect should be captured by a single bias ( $\beta$ ) term.
- **BAO measurement:** To measure the BAO feature in Ly $\alpha$  simulations, the box size should be at least four times larger than the  $100 h^{-1}\cdot\text{Mpc}$  BAO scale [76]. The current Ly $\alpha$  BAO modeling is performed with dark matter simulations. The effect of baryons and Ly $\alpha$  systematics are modeled in post-processing with empirical biases. Here, we wish to perform a "tour de force" of measuring BAO with hydrodynamical simulations to detect baryonic effects at these scales. In contrast, the large scales involved make it tricky to model IGM at sufficient resolution down to the Ly $\alpha$  forest redshift of interest. As discussed in Sec. 3.2.3, correctly modeling the Ly $\alpha$  forest smallest scales requires a resolution

of the order of 30 kpc. This size resolves four times the IGM characteristic scale of the gravitational collapse  $\lambda_{\text{J,IGM}} \sim 100 h^{-1} \cdot \text{kpc}$  [52]. However, as detailed in [52], solving one time this length at  $z = 2$  allows to reach a Ly $\alpha$  estimation accuracy of the 5% order. This level of precision is sufficient in the context of  $P_{3\text{D},\alpha}$  estimation.

### Simulation performed

A first set of simulations in this project was realized on the new GPU partition of the Jean-Zay supercomputer<sup>11</sup> of IDRIS (CNRS).

The initial conditions were generated using the `monofonic` software detailed in Sec. 3.2.2 at  $z_{\text{start}} = 49$ .

We were able to use up to 300 nodes of 4 GPUs, representing almost the entire new partition. We compared the execution of the Nyx GPU version with its CPU implementation by performing simulations of the same resolution with both versions. We showed that Nyx GPU implementation was approximately up to 100 times faster in execution time, considering the number of GPUs and CPUs used. An improvement factor of nearly 2.5 in the calculation time was derived considering the components' energy consumption. The key problem of this type of simulation is the memory usage. As we were limited by the available memory, we could perform the simulations detailed in Tab. 3.1. This simulation set has allowed us to estimate the impacts of the resolution and the box size. Unresolved simulations (resolution higher than  $200 h^{-1} \cdot \text{kpc}$ ) were also performed for testing.

### Nyx improvement

The massive use of a large number of GPU shed light on some Nyx issues and allowed the development team to develop a more stable GPU version. In particular, we highlighted areas of wide-spread shock in the temperature map as represented in Fig. 3.6. This type of shock was also encountered in other studies using the Nyx software [77, 78] and not specific to GPUs. The wide-spread shocks are an interesting phenomenon because of their rarity and the challenge to model them when mapping the dark matter density with the baryon temperature.

As shown in Fig. 3.6, the central region shows temperatures up to  $10^8$  K inside overdensities. Such a temperature is expected in a hydrodynamical simulation and physical. However, for a small number of overdense pixels, the temperature can become unstable and reach non-physical values near  $10^{12}$  K. These numerical instabilities appear more frequently and strongly when using GPUs. After highlighting this phenomenon, the development team corrected it for both CPU and GPU software versions.

### Three dimensional power spectrum estimation

The three-dimensional Ly $\alpha$  forest power spectrum is computed for each simulation, using the `gimlet` software. The  $P_{3\text{D},\alpha}$  is expressed as a function of  $\vec{k}$ . This vector is decomposed between its norm  $k$  and the cosine  $\mu$  of the angle between a reference direction (one of the simulation box axis) and the direction of  $\vec{k}$ . For all simulations, the average transmitted flux fraction  $\bar{F}$  is normalized to

$$\bar{F} = \exp(-\tau_{\text{eff}}) = \exp\left(-0.0025 \times (1+z)^{3.7}\right). \quad (3.21)$$

---

11. <http://www.idris.fr/jean-zay/>

Name	$N_{\text{cell}}$	Size [ $h^{-1}\cdot\text{Mpc}$ ]	Resolution [ $h^{-1}\cdot\text{kpc}$ ]
<i>L500</i>	$2560^3$	500	200
<i>L250</i>	$2560^3$	250	100
<i>L125</i>	$2560^3$	125	50
<i>S250</i>	$1280^3$	250	200
<i>S125</i>	$1280^3$	125	100
<i>S62</i>	$1280^3$	62.5	50
<i>VS125</i>	$640^3$	125	200

TABLE 3.1 – Number of cells, box size, and resolution of simulations used to calculate the three-dimensional power spectrum.

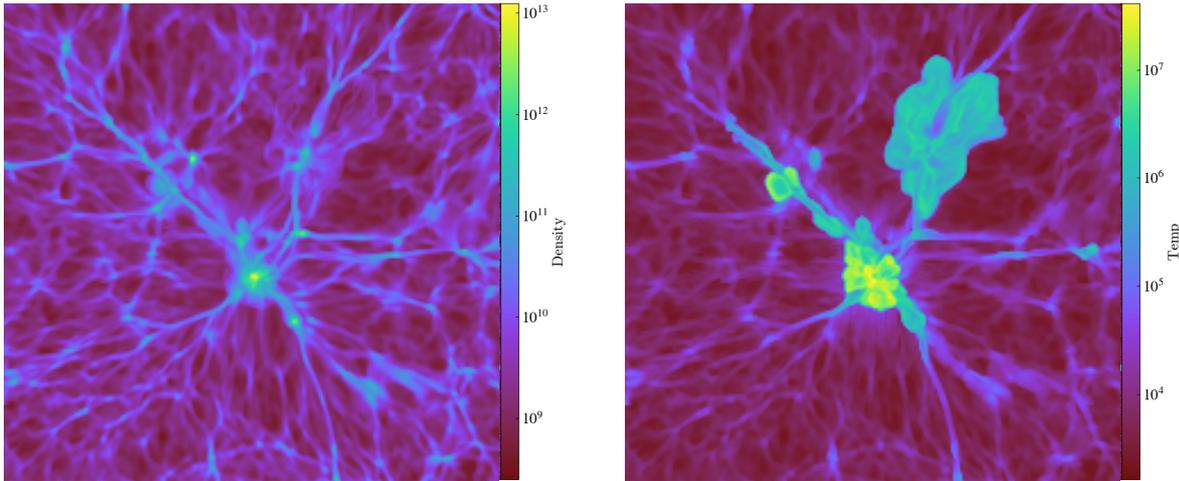


FIGURE 3.6 – Zoom on an overdense region of the simulation which show matter density (left in  $M_{\odot}\cdot\text{Mpc}^{-3}$ ) and temperature (right in  $K$ ). The size of this zoom is about 10Mpc physical at  $z = 2.0$ . This region exhibits high temperature in the middle of the overdensity and a wide-spread shock in a filament. The simulation used to produce this snapshot is not part of the grid given in Tab. 3.1.

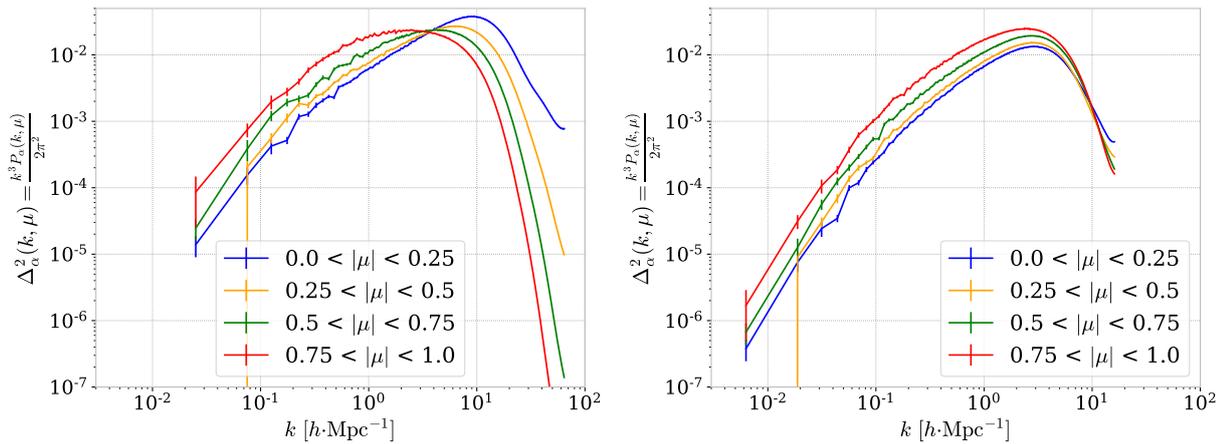


FIGURE 3.7 – Dimensionless three-dimensional power spectrum, as defined by Eqn. 1.71, of the *L125* (left) and *L500* (right) simulations, at  $z = 2.0$ . For both  $P_{3D,\alpha}$ , four  $\mu$  bins with equal size are used.

This normalization follows [63] and is used to compare our simulations with those in this paper and simulations of Sec. 3.2.3.

The three dimensional power spectrum of the largest simulation ( $L500$ ) and the most resolved one ( $L125$ ) are represented in Fig. 3.7. For  $L125$ , the resolution is sufficient (twice the IGM Jeans length) to resolve small scales. We obtain the expected  $P_{3D,\alpha}$  profile as in previous studies [74]. The  $L500$  simulation measures  $P_{3D,\alpha}$  at larger scales ( $k \lesssim 0.02 h\cdot\text{Mpc}^{-1}$ ). However, this simulation is under-resolved and does not model correctly the Jeans smoothing and thermal broadening. Indeed, as shown in previous Nyx studies [52], the resolution of  $L500$  is not sufficient to obtain the right  $P_{3D,\alpha}$  profile.

Physical arguments can explain the shape of  $P_{3D,\alpha}$  for different  $\mu$  bins in  $L125$ . Jeans smoothing for baryons (see Sec. 1.3.3.1) is directly taken into account in our hydrodynamical simulations: the baryonic matter density is smoothed in all directions, which imposes an isotropic cut-off on  $P_{3D,\alpha}$ . The thermal broadening is specific to the Ly $\alpha$  forest and by definition it is visible only along the lines-of-sight. Consequently, this broadening impacts the  $P_{3D,\alpha}$  in an anisotropic way: the effect is maximum for  $|\mu| = 1$  and vanishes transversely ( $|\mu| \sim 0$ ).

As shown by Fig. 3.7 (left) for  $L125$ , both effects impact  $P_{3D,\alpha}$  at small scales ( $k \gtrsim 2 h\cdot\text{Mpc}^{-1}$ ) but the cut-offs related to thermal broadening and Jeans smoothing starts for different  $k$ . Indeed, the thermal broadening suppression starts appearing at scales  $k \sim 2 h\cdot\text{Mpc}^{-1}$ . Conversely, Jeans smoothing impose a power suppression which starts at smaller scales ( $k \sim 10 h\cdot\text{Mpc}^{-1}$ ). The near-transverse  $\mu$  bin ( $|\mu| < 0.25$ ) yields a measurement of Jeans smoothing while the near-radial  $\mu$  bin ( $0.75 < |\mu| < 1$ ) witnesses the combined effect of Jeans and thermal smoothing, and is dominated by thermal broadening. Between both  $\mu$  bins, the transition is progressive.

At larger scales ( $k \lesssim 1 h\cdot\text{Mpc}^{-1}$ ), the difference between  $\mu$  bins is mainly driven by the Kaiser equation 2.40.

### Splicing

As shown in Fig. 3.7,  $L125$  has a sufficient resolution to correctly model  $P_{3D,\alpha}$  [52]. In our study, we wish to keep this resolution while using the large-scale information simulated by  $L500$ . **Splicing** is a numerical trick that utilizes high-resolution simulations to correct the power spectrum of lower resolution simulations with a larger size. This type of technique was used in [73] for the three-dimensional power spectrum estimation, and in [63] to estimate  $P_{1D,\alpha}$ .

Let us assume we have two simulations at hand, with respectively large and small box sizes  $L$  and  $S$ , and small and high resolutions  $Sr$  and  $Hr$ . The splicing method consists in complementing the simulations  $(S,Hr)$  and  $(L,Sr)$  with a "cheap" simulation  $(S,Sr)$  to approximate the three-dimensional power spectrum of a hypothetical  $(L,Hr)$  simulation. This approximated power spectrum, so-called spliced, is defined by using adapted renormalizations for different wavenumber regions. We define the large scales by  $k < k_{\min}$  and the small scales by  $k > k_{\max}$ . The limit wavenumbers are defined a priori by

$$\begin{aligned} k_{\min} &= \frac{2\pi}{S}, \\ k_{\max} &= \frac{k_{\text{Nyq}}}{4} = \frac{\pi Sr}{4L}, \end{aligned} \tag{3.22}$$

where  $k_{\text{Nyq}}$  is the Nyquist frequency of the  $(L,Sr)$  simulation. However, a better recipe for  $k_{\max}$  is to choose it as the last wavenumber for which the  $P_{3D,\alpha}$  difference between  $(L,Sr)$  and  $(S,Sr)$  simulations is lower than 1%. The spliced power spectrum  $P_{\alpha}^{\text{sp}}$  is expressed by

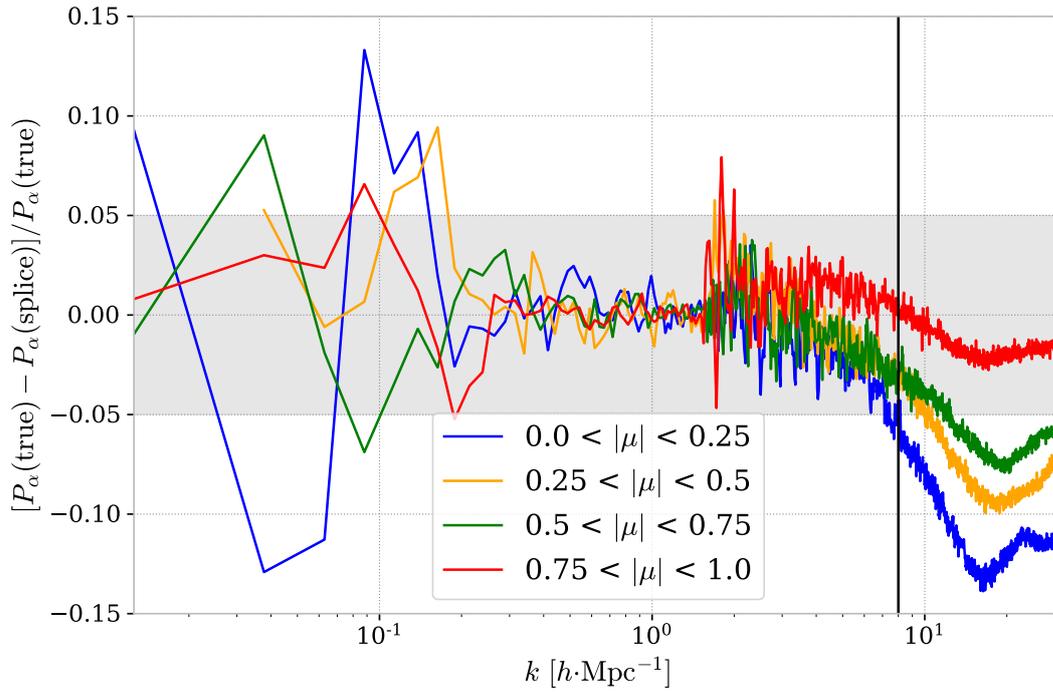


FIGURE 3.8 – Comparison between  $P_{3D,\alpha}$  calculated from the  $L250$  grid and the spliced spectra computed with  $VS125$ ,  $S125$  and  $S250$  simulations. The dark line represents the maximal wavenumber used for the fit. The gray brand corresponds to a 5% relative difference.

$$P_{\alpha}^{\text{sp}}(\vec{k}) = \begin{cases} P_{\alpha}^{(L,\text{Sr})}(\vec{k}) \frac{P_{\alpha}^{(S,\text{Hr})}(k_{\min}, \mu)}{P_{\alpha}^{(S,\text{Sr})}(k_{\min}, \mu)}, & \text{for } |\vec{k}| < k_{\min}, \\ P_{\alpha}^{(L,\text{Sr})}(\vec{k}) \frac{P_{\alpha}^{(S,\text{Hr})}(\vec{k})}{P_{\alpha}^{(S,\text{Sr})}(\vec{k})}, & \text{for } k_{\min} < |\vec{k}| < k_{\max}, \\ P_{\alpha}^{(S,\text{Hr})}(\vec{k}) \frac{P_{\alpha}^{(L,\text{Sr})}(k_{\max}, \mu)}{P_{\alpha}^{(S,\text{Sr})}(k_{\max}, \mu)}, & \text{for } |\vec{k}| > k_{\max}. \end{cases} \quad (3.23)$$

We performed two different splicings with the simulations listed in Tab. 3.1. The first splicing is employed to check the errors introduced by this method. We used a splicing that can be verified, i.e., for which we directly computed the  $P_{3D,\alpha}$  of a (L,Hr) simulation. The box sizes used are  $(L,S) = (250, 125) h^{-1}\cdot\text{Mpc}$  with resolutions  $(\text{Hr},\text{Lr}) = (100, 200) h^{-1}\cdot\text{kpc}$ . The corresponding simulations are  $VS125$ ,  $S125$  and  $S250$ . The difference between the spliced spectra and the  $L250$  three-dimensional power spectrum is shown in Fig. 3.8. For the smallest scales, the bins in  $|\mu| \sim 1$  have a much smaller error than those in  $|\mu| \sim 0$ . It implies that the splicing method is more precise to compute  $P_{1D,\alpha}$ , like in [63]. Considering large scales, the difference is mainly dominated by the errors on the  $P_{3D,\alpha}$  estimation due to the low number of modes. If one requires a 5% precision, the splicing method is precise except for small  $\mu$  and large  $k$ . In our case, we use this test to define the fitting range when a spliced power spectrum is used. We choose a maximum usable wavenumber  $k = 8 h\cdot\text{Mpc}^{-1}$ .

### Three-dimensional power spectrum model

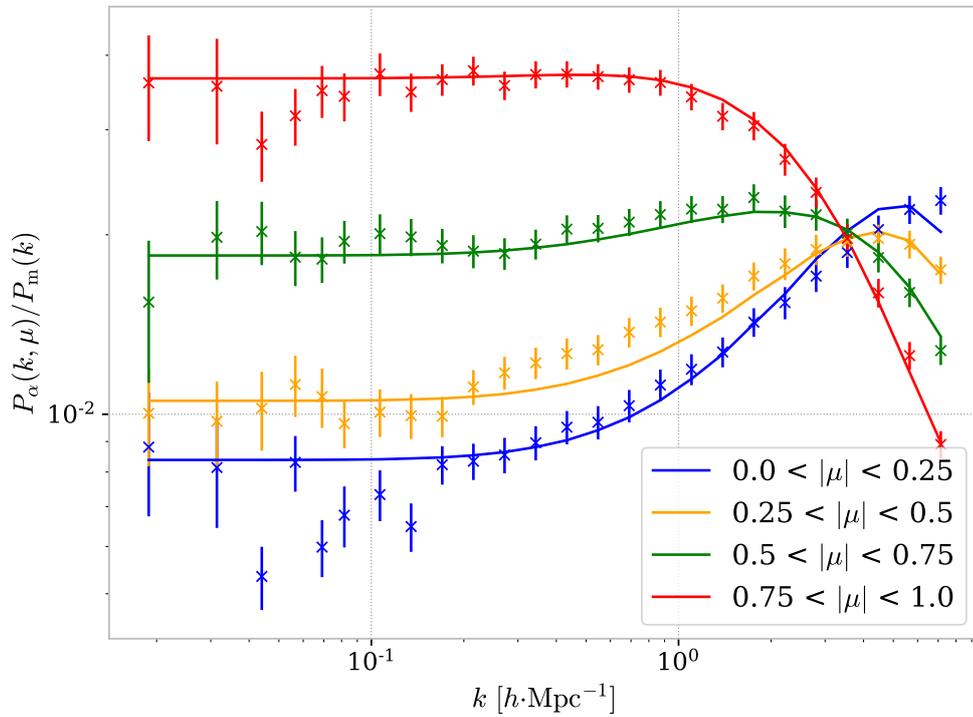


FIGURE 3.9 – (left) Ratio between  $P_{3D,\alpha}$  and the linear matter power spectrum for the spliced simulation used for the fit at  $z = 2.0$ . Both power spectra are computed by splicing the *VS125*, *L125* and *L500* simulations. Each linear matter power spectrum is estimated at the start of the simulation ( $z_{\text{start}} = 49$ ) and normalized to  $z = 2.0$  using CLASS. The estimation from simulations is represented by points and the best-fit model with the  $D_1$  non-linear parameterization by solid lines. The power spectrum crossing at  $k \sim 3 h\cdot\text{Mpc}^{-1}$  is expected and caused by the  $\mu$ -dependent effect of thermal broadening.

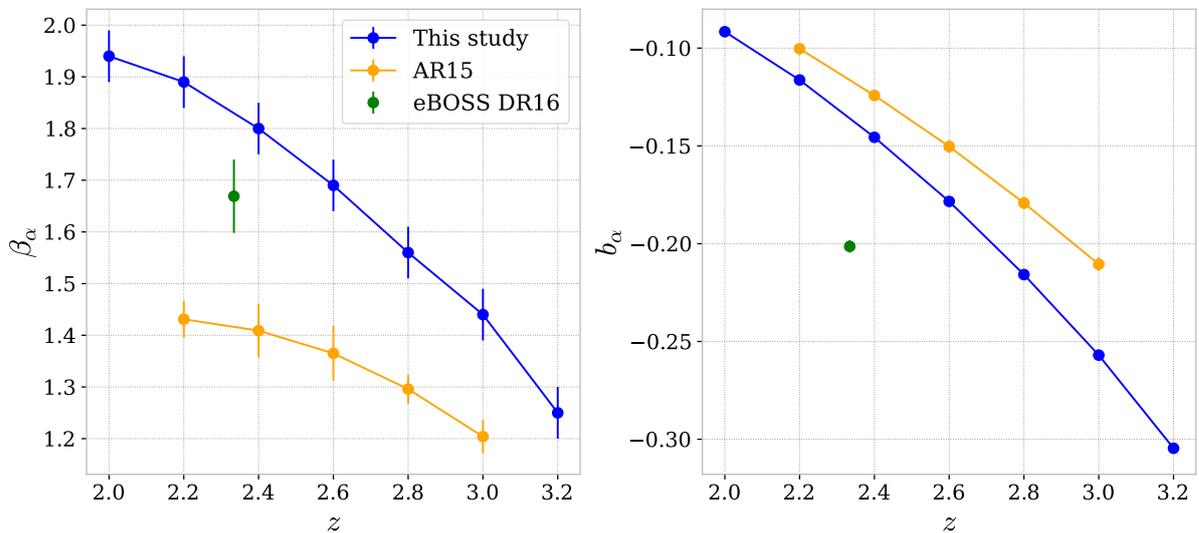


FIGURE 3.10 – Fitted values in blue for the parameters  $b_\alpha$  and  $\beta_\alpha$  on the spliced power spectrum as a function of redshift. The yellow points (AR15) correspond to the prediction from [74] with the same fitting model. The green point is the measurement of these parameters with eBOSS DR16 data [9] using  $\text{Ly}\alpha \times \text{Ly}\alpha$  and  $\text{Ly}\alpha \times \text{QSO}$  large-scale correlations.

We now parameterize  $P_{3D,\alpha}$  along the work of [73, 74]. The principle is to correct Eqn. 2.40 developed in the linear framework, by a non-linear term  $D(k, \mu)$  such that:

$$P_{\text{model},\alpha}(k, \mu) = b_\alpha^2 \left(1 + \beta_\alpha \mu^2\right)^2 P_m(k) D(k, \mu). \quad (3.24)$$

In this equation, all quantities ( $P_m, b_\alpha, \beta_\alpha, D$ ) also depend on redshift. To model the non-linear contribution, a first model was developed in [73] and rewritten in [74] as

$$D_0(k, \mu) = \exp \left\{ \left( \frac{k}{k_{\text{nl}}} \right)^{a_{\text{nl}}} - \left( \frac{k}{k_{\text{p}}} \right)^{a_{\text{p}}} - \left( \frac{k \times \mu}{k_{\text{v}0} (1 + k/k_{\text{v}1})^{a_{\text{v}1}}} \right)^{a_{\text{v}0}} \right\}, \quad (3.25)$$

where  $k_{\text{nl}}$ ,  $a_{\text{nl}}$ ,  $k_{\text{p}}$ ,  $a_{\text{p}}$ ,  $k_{\text{v}0}$ ,  $k_{\text{v}1}$ ,  $a_{\text{v}0}$ , and  $a_{\text{v}1}$  are free parameters. The three corrective terms correspond respectively to the effects of non-linear growth, Jeans smoothing suppression, and the associated effect of thermal broadening and non-linear peculiar velocities along the line-of-sight. A second model developed in [74] uses perturbation theory predictions to reduce the number of free parameters and include the matter power spectrum  $P_m$ :

$$D_1(k, \mu) = \exp \left\{ \left[ q_1 \frac{k^3 P_m(k)}{2\pi^2} + q_2 \left( \frac{k^3 P_m(k)}{2\pi^2} \right)^2 \right] \left[ 1 - \left( \frac{k}{k_{\text{v}}} \right)^{a_{\text{v}}} \mu^{b_{\text{v}}} \right] - \left( \frac{k}{k_{\text{p}}} \right)^2 \right\}, \quad (3.26)$$

where  $q_1$ ,  $q_2$ ,  $k_{\text{v}}$ ,  $a_{\text{v}}$ ,  $b_{\text{v}}$ , and  $k_{\text{p}}$  are free parameters. Here, the non-linear growth corrections at different orders are controlled by the parameters  $q_1$  and  $q_2$ . The Jeans smoothing is ruled by  $k_{\text{p}}$ . Finally, the thermal broadening is handled by  $(k_{\text{v}}, a_{\text{v}}, b_{\text{v}})$ .

### Model fitting

We compute a spliced power spectrum from the simulations *L500*, *L125*, and *VS125*. This splicing takes advantage of our most resolved and largest box size simulations, it corresponds to  $(L, S) = (500, 125) h^{-1} \cdot \text{Mpc}$  with resolutions  $(\text{Hr}, \text{Lr}) = (50, 200) h^{-1} \cdot \text{kpc}$ . The spliced three-dimensional power spectrum, renormalized by the linear matter power spectrum at identical redshift, is shown in Fig. 3.9. The error bars for  $P_{3D,\alpha}$  are computed along the lines of [74]: to avoid to give excessive weight to the smallest scales, we add a parameter  $\epsilon = 0.05$  such that:

$$\sigma_P(k, \mu) = P_\alpha(k, \mu) \left[ \frac{1}{\sqrt{n_\alpha(k, \mu)}} + \epsilon \right], \quad (3.27)$$

where  $n_\alpha(k, \mu)$  is the number of Fourier modes used to compute the  $(k, \mu)$  bin. We fit  $P_{3D,\alpha}$  with the model (3.24, 3.25, 3.26), by  $\chi^2$  minimization using the `iminuit` [79] package, implemented in `lyapower`<sup>1</sup>. The  $\chi^2$  function is:

$$\chi^2 = \sum_{(k, \mu)} \frac{(P_\alpha - P_{\text{model},\alpha})^2}{\sigma_P^2}. \quad (3.28)$$

To mitigate the impact of large-scale cosmic variance in the simulation, we compute  $P_m$  from the simulation itself, at the beginning of each simulation ( $z_{\text{start}} = 49$ ), and rescaled to the considered redshift using `CLASS` software. Since the binning is very different between large and small scales, we have logarithmically rebinned the small scales up to  $k = 0.1 h \cdot \text{Mpc}^{-1}$ .

The main goal of this fit is to improve the measurement of  $b_\alpha$  and  $\beta_\alpha$  parameters. For this purpose, we take advantage of the large scales measured with the spliced  $P_{3D,\alpha}$ . As shown

in Fig. 3.9, the ratio of  $P_{3D,\alpha}$  to the linear matter power spectrum gives a "plateau" at  $k \lesssim 0.2 h\cdot\text{Mpc}^{-1}$  which provides a direct measurement of  $(b_\alpha, \beta_\alpha)$ . The redshift dependence of these linear parameters is given in Fig. 3.10. The predicted values of  $(b_\alpha, \beta_\alpha)$  are compared with hydrodynamical simulation [74] and BAO measurement with eBOSS DR16 data [9]. We note that the values of  $\beta_\alpha$  are higher than simulations in [74], but closer to the eBOSS measurement.

### Summary and prospect

This preliminary study provides very encouraging results. In the near future, our goal is to improve the measurement shown in this section to obtain a robust ab-initio prediction of linear and nonlinear  $\text{Ly}\alpha$  parameters. We are convinced that this will be necessary to interpret  $P_{3D,\alpha}$  measurements performed with high-density  $\text{Ly}\alpha$  surveys such as DESI or WEAVE-QSO.

Furthermore, the simulations developed in this section act as an experimental laboratory accurately representing the three-dimensional distribution of dark matter and baryons. For example, they can be used to test and improve the analytical models used in BAO studies [9] or to interpret the  $\text{Ly}\alpha \times \text{Void}$  cross-correlation which is the subject of Chap. 7.

## Bibliography

- [1] R. E. Angulo and O. Hahn, *Large-scale dark matter simulations*, *arXiv:2112.05165 [astro-ph]* (2021) .
- [2] T. Etourneau, *Les forêts Lyman alpha du relevé eBOSS : comprendre les fonctions de corrélation et les systématiques*, phdthesis, Université Paris-Saclay, Sept., 2020.
- [3] S. Chabanier, *Neutrinos and dark matter cosmology with the Lyman- $\alpha$  forest : the interplay between large-scale evolution and small-scale baryonic physics*, phdthesis, Université Paris-Saclay, Oct., 2020.
- [4] M. Walther, E. Armengaud, C. Ravoux, N. Palanque-Delabrouille, C. Yèche and Z. Lukić, *Simulating intergalactic gas for DESI-like small scale Lyman $\alpha$  forest observations*, *Journal of Cosmology and Astroparticle Physics* **2021** (2021) 059.
- [5] P. Coles and B. Jones, *A lognormal model for the cosmological mass distribution*, *Monthly Notices of the Royal Astronomical Society* **248** (1991) 1.
- [6] H. Bi and A. F. Davidsen, *Evolution of Structure in the Intergalactic Medium and the Nature of the Ly $\alpha$  Forest*, *The Astrophysical Journal* **479** (1997) 523.
- [7] J.-M. L. Goff, C. Magneville, E. Rollinde, S. Peirani, P. Petitjean, C. Pichon et al., *Simulations of BAO reconstruction with a quasar Lyman-alpha survey*, *Astronomy & Astrophysics* **534** (2011) A135.
- [8] T. Etourneau, J. M. L. Goff et al., *Synthetic quasar spectra with 3D Lyman-alpha correlation*, in preparation (2022).
- [9] H. d. M. d. Bourboux, J. Rich, A. Font-Ribera, V. d. S. Agathe, J. Farr, T. Etourneau et al., *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Baryon acoustic oscillations with Lyman- $\alpha$  forests*, *The Astrophysical Journal* **901** (2020) 153.
- [10] J. Farr, A. Font-Ribera, H. d. M. d. Bourboux, A. Muñoz-Gutiérrez, F. J. S. Lopez, A. Pontzen et al., *LyaCoLoRe: Synthetic Datasets for Current and Future Lyman- $\alpha$  Forest BAO Surveys*, *Journal of Cosmology and Astroparticle Physics* **2020** (2020) 068.
- [11] L. Hui and N. Y. Gnedin, *Equation of state of the photoionized intergalactic medium*, *Monthly Notices of the Royal Astronomical Society* **292** (1997) 27.
- [12] S. Chabanier, N. Palanque-Delabrouille, C. Yèche, J.-M. L. Goff, E. Armengaud, J. Bautista et al., *The one-dimensional power spectrum from the SDSS DR14 Ly $\alpha$  forests*, *Journal of Cosmology and Astroparticle Physics* **2019** (2019) 017.
- [13] U. Seljak, *Bias, redshift space distortions and primordial nongaussianity of nonlinear transformations: application to Lyman alpha forest*, *Journal of Cosmology and Astroparticle Physics* **2012** (2012) 004.
- [14] N. G. Karaçaylı, A. Font-Ribera and N. Padmanabhan, *Optimal 1D Ly- $\alpha$  Forest Power Spectrum Estimation I: DESI-Lite Spectra*, *Monthly Notices of the Royal Astronomical Society* **497** (2020) 4742.

- [15] N. Palanque-Delabrouille, C. Yèche, A. Borde, J.-M. L. Goff, G. Rossi, M. Viel et al., *The one-dimensional Ly-alpha forest power spectrum from BOSS*, *Astronomy & Astrophysics* **559** (2013) A85.
- [16] N. Palanque-Delabrouille, C. Magneville, C. Yèche, I. Pâris, P. Petitjean, E. Burtin et al., *The extended Baryon Oscillation Spectroscopic Survey: Variability selection and quasar luminosity function*, *Astronomy & Astrophysics* **587** (2016) A41.
- [17] DESI collaboration, *The DESI Experiment Part I: Science, Targeting, and Survey Design*, *arXiv:1611.00036 [astro-ph]* (2016) .
- [18] A. González-Morales et al., *DESI Lyman-alpha synthetic spectra*, in preparation (2022).
- [19] I. D. McGreer, L. Jiang, X. Fan, G. T. Richards, M. A. Strauss, N. P. Ross et al., *The  $z = 5$  Quasar Luminosity Function from SDSS Stripe 82*, *The Astrophysical Journal* **768** (2013) 105.
- [20] S. Chabanier, T. Etourneau, J.-M. L. Goff, J. Rich, J. Stermer, B. Abolfathi et al., *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: The Damped Lyman- $\alpha$  systems Catalog*, *arXiv:2107.09612 [astro-ph, physics:physics]* (2021) .
- [21] W. H. Press and P. Schechter, *Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation*, *The Astrophysical Journal* **187** (1974) 425.
- [22] D. Potter, J. Stadel and R. Teyssier, *PKDGRAV3: Beyond Trillion Particle Cosmological Simulations for the Next Era of Galaxy Surveys*, *arXiv:1609.08621 [astro-ph]* (2016) .
- [23] Y. Dubois, S. Peirani, C. Pichon, J. Devriendt, R. Gavazzi, C. Welker et al., *The Horizon-AGN simulation: morphological diversity of galaxies promoted by AGN feedback*, *Monthly Notices of the Royal Astronomical Society* **463** (2016) 3948.
- [24] N. A. Maksimova, L. H. Garrison, D. J. Eisenstein, B. Hadzhiyska, S. Bose and T. P. Satterthwaite, *AbacusSummit: a massive set of high-accuracy, high-resolution N-body simulations*, *Monthly Notices of the Royal Astronomical Society* **508** (2021) 4017.
- [25] K. Heitmann, H. Finkel, A. Pope, V. Morozov, N. Frontiere, S. Habib et al., *The Outer Rim Simulation: A Path to Many-Core Supercomputers*, *The Astrophysical Journal Supplement Series* **245** (2019) 16.
- [26] M. Vogelsberger, S. D. M. White, A. Helmi and V. Springel, *The fine-grained phase-space structure of cold dark matter haloes*, *Monthly Notices of the Royal Astronomical Society* **385** (2008) 236.
- [27] S. D. M. White and M. Vogelsberger, *Dark matter caustics*, *Monthly Notices of the Royal Astronomical Society* **392** (2009) 281.
- [28] T. Abel, O. Hahn and R. Kaehler, *Tracing the Dark Matter Sheet in Phase Space*, *Monthly Notices of the Royal Astronomical Society* **427** (2012) 61.
- [29] K. Yoshikawa, N. Yoshida and M. Umemura, *Direct Integration of the Collisionless Boltzmann Equation in Six-dimensional Phase Space: Self-gravitating Systems*, *The Astrophysical Journal* **762** (2013) 116.

- [30] S. Tanaka, K. Yoshikawa, T. Minoshima and N. Yoshida, *Multidimensional Vlasov-Poisson Simulations with High-order Monotonicity- and Positivity-preserving Schemes*, *The Astrophysical Journal* **849** (2017) 76.
- [31] K. Yoshikawa, S. Tanaka, N. Yoshida and S. Saito, *Cosmological Vlasov-Poisson Simulations of Structure Formation with Relic Neutrinos: Nonlinear Clustering and the Neutrino Mass*, *The Astrophysical Journal* **904** (2020) 159.
- [32] K. Yoshikawa, S. Tanaka and N. Yoshida, *A 400 Trillion-Grid Vlasov Simulation on Fugaku Supercomputer: Large-Scale Distribution of Cosmic Relic Neutrinos in a Six-dimensional Phase Space*, *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (2021) 1.
- [33] J. Dakin, J. Brandbyge, S. Hannestad, T. HaugbØlle and T. Tram,  $\nu$ CO N CEPT: *cosmological neutrino simulations from the non-linear Boltzmann hierarchy*, *Journal of Cosmology and Astroparticle Physics* **2019** (2019) 052.
- [34] T. Tram, J. Brandbyge, J. Dakin and S. Hannestad, *Fully relativistic treatment of light neutrinos in N-body simulations*, *Journal of Cosmology and Astroparticle Physics* **2019** (2019) 022.
- [35] C. Partmann, C. Fidler, C. Rampf and O. Hahn, *Fast simulations of cosmic large-scale structure with massive neutrinos*, *Journal of Cosmology and Astroparticle Physics* **2020** (2020) 018.
- [36] J. Brandbyge and S. Hannestad, *Grid based linear neutrino perturbations in cosmological N-body simulations*, *Journal of Cosmology and Astroparticle Physics* **2009** (2009) 002.
- [37] S. Bird, M. Viel and M. G. Haehnelt, *Massive Neutrinos and the Non-linear Matter Power Spectrum*, *Monthly Notices of the Royal Astronomical Society* **420** (2012) 2551.
- [38] Y. Ali-Haïmoud and S. Bird, *An efficient implementation of massive neutrinos in non-linear structure formation simulations*, *Monthly Notices of the Royal Astronomical Society* **428** (2013) 3375.
- [39] A. Banerjee, D. Powell, T. Abel and F. Villaescusa-Navarro, *Reducing noise in cosmological N-body simulations with neutrinos*, *Journal of Cosmology and Astroparticle Physics* **2018** (2018) 028.
- [40] J. Brandbyge, S. Hannestad and T. Tram, *Momentum space sampling of neutrinos in N-body simulations*, *Journal of Cosmology and Astroparticle Physics* **2019** (2019) 047.
- [41] S. Bird, Y. Ali-Haïmoud, Y. Feng and J. Liu, *An efficient and accurate hybrid method for simulating non-linear neutrino structure*, *Monthly Notices of the Royal Astronomical Society* **481** (2018) 1486.
- [42] M. Zennaro, R. E. Angulo, G. Aricò, S. Contreras and M. Pellejero-Ibáñez, *How to add massive neutrinos to your  $\Lambda$ CDM simulation – extending cosmology rescaling algorithms*, *Monthly Notices of the Royal Astronomical Society* **489** (2019) 5938.

- [43] C. Pedersen, A. Font-Ribera, K. K. Rogers, P. McDonald, H. V. Peiris, A. Pontzen et al., *An emulator for the Lyman- $\alpha$  forest in beyond- $\Lambda$ CDM cosmologies*, *arXiv:2011.15127 [astro-ph]* (2020) .
- [44] C. Pedersen, A. Font-Ribera, T. D. Kitching, P. McDonald, S. Bird, A. Slosar et al., *Massive neutrinos and degeneracies in Lyman-alpha forest simulations*, *arXiv:1911.09596 [astro-ph]* (2020) .
- [45] J. Lesgourgues, *The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview*, *arXiv:1104.2932 [astro-ph]* (2011) .
- [46] D. Blas, J. Lesgourgues and T. Tram, *The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes*, *Journal of Cosmology and Astroparticle Physics* **2011** (2011) 034.
- [47] M. Crocce, S. Pueblas and R. Scoccimarro, *2LPTIC: 2nd-order Lagrangian Perturbation Theory Initial Conditions*, *Astrophysics Source Code Library* (2012) ascl:1201.005.
- [48] O. Hahn, C. Rampf and C. Uhlemann, *Higher-order initial conditions for mixed baryon-CDM simulations*, *arXiv:2008.09124 [astro-ph]* (2020) .
- [49] M. Michaux, O. Hahn, C. Rampf and R. E. Angulo, *Accurate initial conditions for cosmological N-body simulations: Minimizing truncation and discreteness errors*, *arXiv:2008.09588 [astro-ph]* (2020) .
- [50] M. A. Fernandez, S. Bird and P. Upton Sanderbeck, *Effect of separate initial conditions on the Lyman- $\alpha$  forest in simulations*, *Monthly Notices of the Royal Astronomical Society* **503** (2021) 1668.
- [51] A. S. Almgren, J. B. Bell, M. J. Lijewski, Z. Lukić and E. Van Andel, *Nyx: A MASSIVELY PARALLEL AMR CODE FOR COMPUTATIONAL COSMOLOGY*, *The Astrophysical Journal* **765** (2013) 39.
- [52] Z. Lukić, C. Stark, P. Nugent, M. White, A. Meiksin and A. Almgren, *The Lyman- $\alpha$  forest in optically-thin hydrodynamical simulations*, *arXiv:1406.6361 [astro-ph]* (2015) .
- [53] S. Kaviraj, C. Laigle, T. Kimm, J. E. G. Devriendt, Y. Dubois, C. Pichon et al., *The Horizon-AGN simulation: evolution of galaxy properties over cosmic time*, *Monthly Notices of the Royal Astronomical Society* (2017) stx126.
- [54] N. E. Chisari, M. L. A. Richardson, J. Devriendt, Y. Dubois, A. Schneider, A. M. C. Le Brun et al., *The impact of baryons on the matter power spectrum from the Horizon-AGN cosmological hydrodynamical simulation*, *Monthly Notices of the Royal Astronomical Society* **480** (2018) 3962.
- [55] S. Chabanier, F. Bournaud, Y. Dubois, N. Palanque-Delabrouille, C. Yèche, E. Armengaud et al., *The impact of AGN feedback on the 1D power spectra from the Ly $\alpha$  forest using the Horizon-AGN suite of simulations*, *Monthly Notices of the Royal Astronomical Society* **495** (2020) 1825.

- [56] S. Gontcho A Gontcho, J. Miralda-Escudé and N. G. Busca, *On the effect of the ionizing background on the Ly $\alpha$  forest autocorrelation function*, *Monthly Notices of the Royal Astronomical Society* **442** (2014) 187.
- [57] D. Aubert and R. Teyssier, *Reionization simulations powered by GPUs I: the structure of the Ultraviolet radiation field*, *The Astrophysical Journal* **724** (2010) 244.
- [58] P. Ocvirk, D. Aubert, J. G. Sorce, P. R. Shapiro, N. Deparis, T. Dawoodbhoy et al., *Cosmic Dawn II (CoDa II): a new radiation-hydrodynamics simulation of the self-consistent coupling of galaxy formation and reionization*, *Monthly Notices of the Royal Astronomical Society* **496** (2020) 4087.
- [59] J. S. W. Lewis, P. Ocvirk, D. Aubert, J. G. Sorce, P. R. Shapiro, N. Deparis et al., *Galactic ionizing photon budget during the epoch of reionization in the Cosmic Dawn II simulation*, *Monthly Notices of the Royal Astronomical Society* **496** (2020) 4342.
- [60] P. Montero-Camacho, Y. Liu and Y. Mao, *Separating the memory of reionization from cosmology in the Ly $\alpha$  forest power spectrum at the post-reionization era*, July, 2022. 10.48550/arXiv.2207.09005.
- [61] B. Friesen, A. Almgren, Z. Lukić, G. Weber, D. Morozov, V. Beckner et al., *In situ and in-transit analysis of cosmological simulations*, *Computational Astrophysics and Cosmology* **3** (2016) 4.
- [62] M. Viel, M. G. Haehnelt and A. Lewis, *The Lyman-alpha forest and WMAP year three*, *Monthly Notices of the Royal Astronomical Society: Letters* **370** (2006) L51.
- [63] A. Borde, N. Palanque-Delabrouille, G. Rossi, M. Viel, J. Bolton, C. Yèche et al., *New approach for precise computation of Lyman-alpha forest power spectrum with hydrodynamical simulations*, *Journal of Cosmology and Astroparticle Physics* **2014** (2014) 005.
- [64] M. Walther, J. Oñorbe, J. F. Hennawi and Z. Lukić, *New Constraints on IGM Thermal Evolution from the Ly $\alpha$  Forest Power Spectrum*, *The Astrophysical Journal* **872** (2019) 13.
- [65] S. Bird, K. K. Rogers, H. V. Peiris, L. Verde, A. Font-Ribera and A. Pontzen, *An Emulator for the Lyman-alpha Forest*, *Journal of Cosmology and Astroparticle Physics* **2019** (2019) 050.
- [66] K. K. Rogers, H. V. Peiris, A. Pontzen, S. Bird, L. Verde and A. Font-Ribera, *Bayesian emulator optimisation for cosmology: application to the Lyman-alpha forest*, *Journal of Cosmology and Astroparticle Physics* **2019** (2019) 031.
- [67] M. Viel, J. Weller and M. Haehnelt, *Constraints on the Primordial Power Spectrum from High Resolution Lyman-alpha Forest Spectra and WMAP*, *Monthly Notices of the Royal Astronomical Society* **355** (2004) L23.
- [68] G. D. Becker, J. S. Bolton, M. G. Haehnelt and W. L. W. Sargent, *Detection of Extended He II Reionization in the Temperature Evolution of the Intergalactic Medium*, *Monthly Notices of the Royal Astronomical Society* **410** (2011) 1096.

- [69] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astronomy & Astrophysics* **641** (2020) A6.
- [70] S. Ambikasaran, D. Foreman-Mackey, L. Greengard, D. W. Hogg and M. O’Neil, *Fast Direct Methods for Gaussian Processes*, *arXiv:1403.6015 [astro-ph, stat]* (2015) .
- [71] N. Palanque-Delabrouille, C. Yèche, N. Schöneberg, J. Lesgourgues, M. Walther, S. Chabanier et al., *Hints, neutrino bounds and WDM constraints from SDSS DR14 Lyman- $\alpha$  and Planck full-survey data*, *Journal of Cosmology and Astroparticle Physics* **2020** (2020) 038.
- [72] A. Font-Ribera, P. McDonald and A. Slosar, *How to estimate the 3D power spectrum of the Lyman- $\alpha$  forest*, *Journal of Cosmology and Astroparticle Physics* **2018** (2018) 003.
- [73] P. McDonald, *Toward a Measurement of the Cosmological Geometry at  $z \sim 2$ : Predicting Ly $\alpha$  Forest Correlation in Three Dimensions and the Potential of Future Data Sets*, *The Astrophysical Journal* **585** (2003) 34.
- [74] A. Arinyo-i Prats, J. Miralda-Escudé, M. Viel and R. Cen, *The non-linear power spectrum of the Lyman alpha forest*, *Journal of Cosmology and Astroparticle Physics* **2015** (2015) 017.
- [75] K. K. Rogers, S. Bird, H. V. Peiris, A. Pontzen, A. Font-Ribera and B. Leistedt, *Correlations in the three-dimensional Lyman-alpha forest contaminated by high column density absorbers*, *Monthly Notices of the Royal Astronomical Society* **476** (2018) 3716.
- [76] A. Klypin and F. Prada, *Effects of long-wavelength fluctuations in large galaxy surveys*, *Monthly Notices of the Royal Astronomical Society* **489** (2019) 1684.
- [77] P. Harrington, M. Mustafa, M. Dornfest, B. Horowitz and Z. Lukić, *Fast, high-fidelity Lyman  $\alpha$  forests with convolutional neural networks*, *arXiv:2106.12662 [astro-ph]* (2021) .
- [78] B. Horowitz, M. Dornfest, Z. Lukić and P. Harrington, *HyPhy: Deep Generative Conditional Posterior Mapping of Hydrodynamical Physics*, *arXiv:2106.12675 [astro-ph]* (2021) .
- [79] H. Dembinski, P. Ongmongkolkul, C. Deil, H. Schreiner, M. Feickert, Andrew et al., *scikit-hep/iminuit: v2.11.0*, Mar., 2022. 10.5281/zenodo.6387620.



# 4

---

## The SDSS and DESI multi-object spectrograph surveys

---

*“Les télescopes, ces cathédrales des temps modernes qui recueillent la lumière du cosmos, sont de véritables machines à remonter le temps.”*

– Trinh Xuan Thuan, *Le Cosmos et le Lotus*, 2011

**Contents**

---

<b>4.1 Galaxy surveys</b> . . . . .	<b>130</b>
4.1.1 Generalities . . . . .	130
4.1.2 Properties of observed spectra . . . . .	132
<b>4.2 Sloan Digital Sky Survey</b> . . . . .	<b>133</b>
4.2.1 History of SDSS . . . . .	133
4.2.2 Instrument description . . . . .	134
4.2.3 Spectroscopic pipeline . . . . .	136
<b>4.3 Dark Energy Spectroscopic Instrument</b> . . . . .	<b>138</b>
4.3.1 Survey overview . . . . .	138
4.3.2 Instrument description . . . . .	142
4.3.2.1 Telescope . . . . .	142
4.3.2.2 Spectrographs . . . . .	146
4.3.3 Observing with DESI . . . . .	147
4.3.4 DESI spectroscopic pipeline . . . . .	150
4.3.4.1 Spectroscopic extraction . . . . .	150
4.3.4.2 Evaluating the noise . . . . .	153
4.3.4.3 Classifying spectra and measuring redshifts . . . . .	156
<b>Bibliography</b> . . . . .	<b>157</b>

---

THE first astronomical observation using an instrument was made by Galileo Galilei at the very beginning of the XVII<sup>th</sup> century. Between Galilei's first astronomical scope and today, the complexity of observation systems has followed an exponential evolution. To give an example, as I write this thesis, the James Webb Space Telescope (JWST), the most complex and expensive telescope ever created, has deployed its 6.5 m primary mirror at a Lagrange point of our Solar System.

The use of these increasingly complex systems has allowed to test and improve our understanding of the Universe through the models exposed in Chap. 1. The increase in the size of the telescope optics allows us to observe objects of fainter and fainter luminosity and thus more and more distant objects. Improving technology in this field is our best way to "go back in time".

Astronomers also aim to perform observations by covering as much of the electromagnetic spectrum as possible. The first optical sensor used by Galilei, his eye, allowed him to see the Universe only through the visible spectrum. Recent telescopes on the ground or in space observe the Universe from radio waves to the gamma rays thanks to the improvement of electromagnetic sensors. Furthermore, astronomers increased their observation capacity by detecting astroparticles other than photons, like neutrinos or high-energy protons. The very recent measurement of gravitational waves adds another observation window of the Universe.

In addition to the improvement of the technologies used to collect the light from celestial bodies, the data processing capacities have drastically increased. Although it is very difficult to criticize Galilei's brain, it has been far surpassed by current supercomputers at least for the task of computing and data processing. Today, interpreting data from the largest telescopes is impossible without the use of a proper pipeline. The improvement of processors and information systems is thus equally necessary to continue unravelling the mystery of the Universe.

During my thesis, I had the chance to work on data from two complex telescopes whose common goal is to trace the distribution of matter in the Universe. By doing so, they allow answering the fundamental questions expressed in Chap. 1. Both telescopes have the ability to measure the optical spectra of several celestial objects simultaneously. The description of this type of system called multi-object spectrograph (MOS) and the data they provide is given in Sec. 4.1. The Sloan Digital Sky Survey (SDSS), the first survey I worked on is described in Sec. 4.2. I also analyzed the first spectra of the recent Dark Energy Spectroscopic Instrument (DESI), which will be described in Sec. 4.3.

## 4.1 Galaxy surveys

### 4.1.1 Generalities

As mentioned in the Sec. 1.4, large galaxy surveys are divided into two types: **photometric surveys** which observe a given footprint with several filters and **spectroscopic surveys** which measure the spectra of extra-galactic objects.

These two types of observations are intended for different goals. Photometric surveys observe all objects in a field with a magnitude below a certain value. However, the determination of the objects' redshift is imprecise. This is due to the fact that the flux is only known as an integral over some range in wavelength as shown in Fig. 4.1 (left). On the other hand, as the measurement of a spectrum requires a longer exposure time, spectroscopic surveys accurately measure the redshifts of a limited number of objects. The choice of a survey for a certain scientific purpose must therefore take into account this trade-off between number of objects and redshift determination.

Spectroscopic surveys also make it possible to carry out studies directly related to the spectra such as the Ly $\alpha$  forest detailed in Chap. 2.

There are several ways to measure galaxy spectra. The set of spectra of all objects observed is a three dimensional data: two spatial coordinates to locate the object in the sky and the spectrum of each object as a function of the wavelength. A schematic representation of this three dimensional data cube is given in Fig. 4.1 (right). The main difficulty in this measurement is to transform this three-dimensional data into a signal measured by a two dimensional detector (usually a CCD camera). This can be achieved by several methods: aperture spectroscopy, slitless spectroscopy or integral field spectroscopy. In this chapter, we will describe only aperture spectroscopy. We focus on the particular case where these measurements are multiplexed by the use of optical fibers. This type of measurement is performed by an instrument called **multi-object spectrograph (MOS)**.

A MOS is generally composed of the following elements:

- **Telescope:** The optical system of the telescope concentrates the light flux from a certain field of view in the sky.
- **Optical fibers:** The fibers select an aperture targeted on the focal plane and redirect the flux towards a spectrograph.
- **Spectrograph:** It decomposes the light from fibers into several wavelengths thanks to a dispersive optical system. This dispersed light is then captured by cameras usually equipped with a charge-coupled device (CCD). A spectrograph is used to analyze the light from several fibers distributed along a CCD direction.

As the number of fibers is limited, one must select the object from which to obtain spectra. For this reason, all MOS surveys require a **target selection** step. This is done from photometric surveys as represented on Fig. 4.1 (right). Targets are selected from the measured magnitudes, colors, spatial extension, and time variability of objects identified from these photometric surveys.

The idea of using MOS to perform galaxy redshift surveys is rather recent in the history of science. The first surveys started in the 1980s. Since then, this type of survey has become increasingly used because of its capacity to map the matter density of the Universe. The number of spectra observed by MOS approximately follows a Moore's law as shown in Fig. 4.2. In

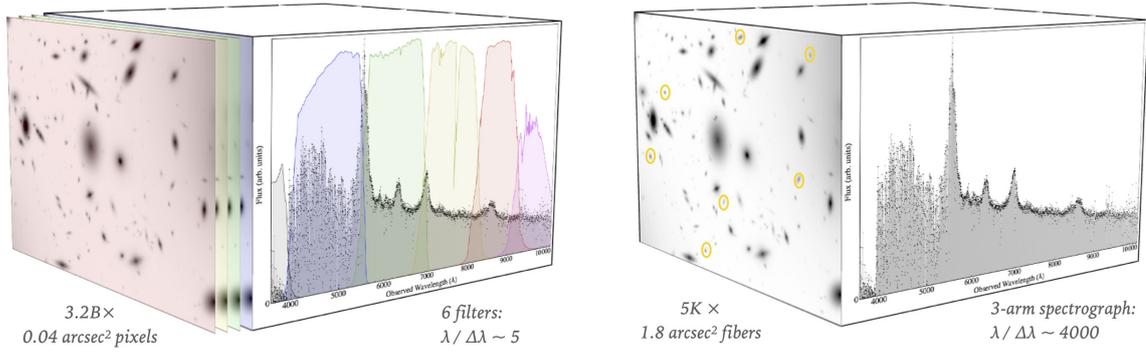


FIGURE 4.1 – (left) Representation of the data product given by a photometric survey. The sky is globally observed with specific filters which give integrated information with respect to wavelength. The numbers reported on this figure corresponds to those of the Rubin observatory [1]. (right) Representation of the data product given by a spectroscopic survey. Based on photometric imaging, some objects are targeted. Their position in the sky, usually given by (RA,DEC) angles, constitute the first two dimensions of the data cube. The spectrum associated to each target is then measured and gives the third dimension of the data cube. The position of a data cube pixel can be expressed in observational coordinates (RA,DEC, $z$ ) or in cartesian coordinates ( $x,y,z$ ) when a cosmological model is used to calculate them. The numbers reported on this figure corresponds to DESI and are detailed in the Sec. 4.3. Credits: DESI key figures for external communication.

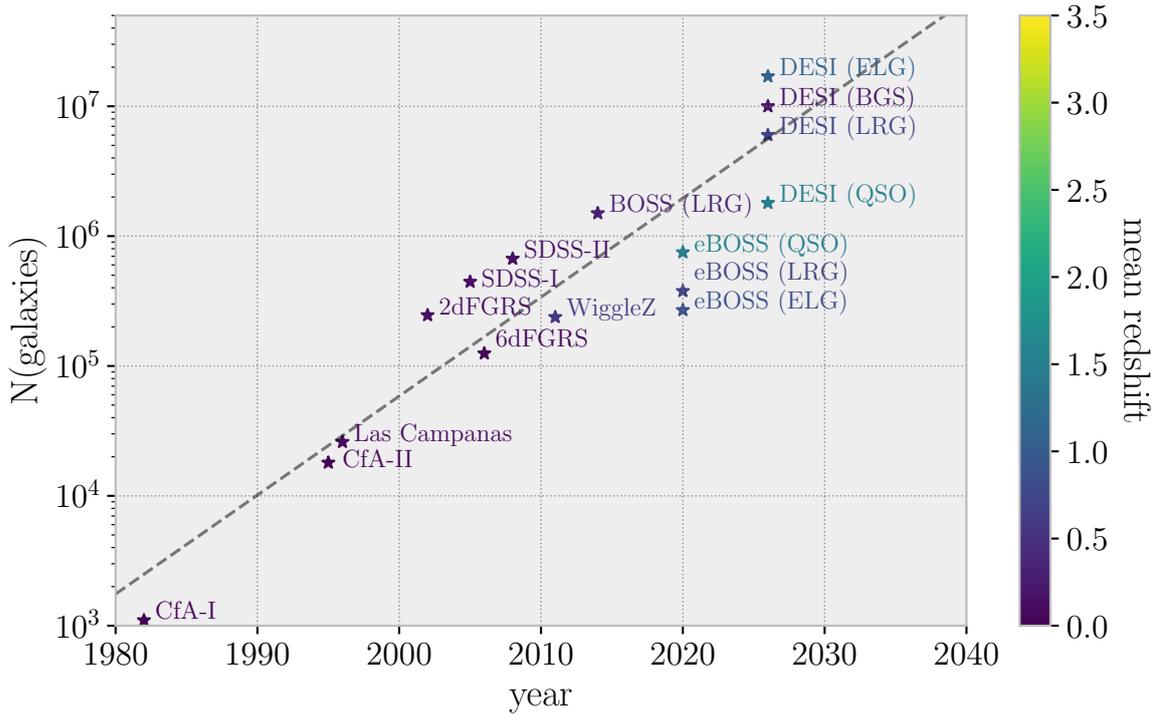


FIGURE 4.2 – Number of extra-galactic objects whose spectra were measured by different galaxy surveys. This number is represented as a function of date and follows approximately a Moore's law represented in dotted line. For DESI, the number of spectra is the one expected after five years. Credits: E. Chaussidon.

addition, as indicated by the color axis representing the average redshift of a group of observed objects, these surveys aim at observing increasingly large redshifts.

## 4.1.2 Properties of observed spectra

An example of spectrum is given in Fig. 2.3. Here, we define important properties of measured spectra.

A spectrum is a flux profile  $f(\lambda)$  (in blue in Fig. 2.3) expressed as a function of the wavelength  $\lambda$ . The units are usually  $10^{-17} \cdot \text{erg} \cdot \text{s}^{-1} \cdot \text{cm}^{-2} \cdot \text{\AA}^{-1}$  for  $f(\lambda)$  and  $\text{\AA}$  for  $\lambda$ . To each spectrum, we associate a noise  $\sigma_f(\lambda)$  (in orange in Fig. 2.3). This noise can be estimated by modeling, by direct measurement or by statistical methods. We will come back to this point in more detail in the case of DESI in the Sec. 4.3.4.2.

A spectrum can be binned in wavelength in a regular way (linearly, logarithmically) or irregularly. We define the size of a pixel as the separation between two successive pixels  $\Delta\lambda_{\text{pix}}$ .

The **signal-to-noise ratio (SNR)** is an important quantity to characterize the quality of a spectrum. The SNR per pixel is given by  $\text{SNR}(\lambda) = f(\lambda)/\sigma_f(\lambda)$ . It is also possible to obtain the SNR per Angstrom by multiplying the SNR per pixel by the separation between each pixel  $\Delta\lambda_{\text{pix}}$ .

A spectrum is characterized by its **spectral resolution**  $R(\lambda) = \frac{\lambda}{\Delta\lambda(\lambda)}$ , where  $\Delta\lambda(\lambda)$  corresponds to the resolution power of the spectrograph. A larger spectral resolution provides more detailed information on narrow spectral features such as emission and absorption lines. However, for an object of given brightness, it requires a longer observation time to have the same SNR per pixel.

## 4.2 Sloan Digital Sky Survey

The Sloan Digital Sky Survey (SDSS) was imagined in the 1980s by Jim Gunn, Rich Kron and Donald York [2], and partly financed by the Sloan foundation. This survey make use of a telescope located at the Apache Point Observatory in New-Mexico. The telescope's optical system was used to run both photometric and spectroscopic surveys, by switching at the focal point an imaging camera or a fiber plate connected to two spectrographs. The same telescope was therefore used for target selection and measurement of the corresponding spectra.

For cosmology, the main goal of this survey was to obtain a better understanding of the large scale structures of the Universe. After the discovery of the acceleration of expansion in 1998, the determination of the properties and nature of dark energy using BAO also became a key objective. The SDSS project has been carried out in four phases and is currently in its fifth phase. Each phase lasts approximately 5 years. In this thesis, the data used are from the last two completed phases of this survey.

The description of SDSS given in this section is based on the reference articles [2, 3, 4]. A review of the history of the different phases of SDSS and its objectives is given in Sec. 4.2.1. A description of the instrument dedicated to the SDSS project and its improvements is given in Sec. 4.2.2. Finally, Sec. 4.2.3 describes the spectroscopic data processing pipeline.

### 4.2.1 History of SDSS

#### SDSS-I/II

The first light of the SDSS telescope took place in 1998. The observing phase SDSS-I lasted between 2000 and 2005. This first phase provided a multi-band photometric survey of approximately  $8000 \text{ deg}^2$  and a spectroscopic survey of galaxies and QSOs over  $5700 \text{ deg}^2$ .

During its second phase SDSS-II between 2005 and 2008, the survey was divided between the continuation of its first phase and two extension projects. The pursuit of SDSS-I aimed to increase the footprint of the photometric surveys up to  $11500 \text{ deg}^2$  and spectroscopic surveys of  $7500 \text{ deg}^2$ . It is in this project, by the observation of particular galaxies called luminous red galaxies (LRGs), that the first detection of the BAO peak was achieved [5]. The two extension projects concern Milky Way observations (SEGUE), and the detection of type Ia Supernovae with spectroscopic follow-up (Sloan Supernovae Survey).

#### SDSS-III

The third phase SDSS-III [6] lasted between 2008 and 2014. It included one extragalactic survey and three surveys related to the observation of the Milky Way. The spectroscopy system was improved before the beginning of this new phase.

The **Baryon Oscillation Spectroscopic Survey (BOSS)** [7] is an extension of the spectroscopic surveys of galaxies from the first phases. The increased statistics of the LRG sample allowed to improve the measurement of the BAO peak. During this phase, many QSOs with a redshift  $z > 2.0$  exhibiting a Ly $\alpha$  forest were observed. A sufficient number of QSOs could be observed during this phase to obtain the first detection of the BAO peak using Ly $\alpha$  forest [8]. In total, the spectra of more than 1.5 million LRGs and 150,000 high-z QSOs were observed during this survey, over a  $10,000 \text{ deg}^2$  footprint.

During this phase, three projects concerning the observation of the Milky Way were carried out. SEGUE-2 in the continuation of its first phase measured stellar spectra inside the Milky

Way. The APOGEE survey measured the peculiar velocities of red giant stars in different regions of our galaxy. Finally, the MARVELS project observed multiple times specific stars of the Milky Way in order to measure their radial velocity, and to potentially detect the presence of exoplanets.

### SDSS-IV

The fourth phase SDSS-IV [9] is the last completed phase. Observations lasted from 2014 to 2019. This phase was separated into three separate surveys. The last data release is detailed in [10].

The **extended Baryon Oscillation Spectroscopic Survey (eBOSS)** [11] is the SDSS survey used in this thesis. Its purpose was to continue the observations made by BOSS. In addition to increasing the statistics of each tracer (nearby galaxies and LRG), this survey made possible the first detection of the BAO peak using QSOs as point tracers at redshift  $z < 2.0$  [12]. Finally, a new type of faint target called emission line galaxies (ELGs) and characterized by the presence of a pronounced OII doublet were observed. This also led to the first tentative measurement of the BAO peak with this tracer [13]. These observations constitute a pathfinder survey for DESI. Clustering measurements with all types of tracer (nearby galaxies, LRG, ELG, QSO and Ly $\alpha$  forest) provided a measurement of the BAO peak (illustrated in Fig. 1.5) and of RSD over a large redshift range. They led to strong constraints on the  $\Lambda$ CDM model and on its possible extensions (neutrinos, dark energy models, modified gravity) [14]. The Ly $\alpha$  forest measured by the BOSS and eBOSS surveys are used in Chap. 6 and 7.

In parallel, the APOGEE-2 survey extended the observation of the Milky Way made in the previous phase. Finally, the MaNGA survey focused on the observation of galaxies very close to the Milky Way, to better understand the history of nearby galaxies by measuring their composition.

### SDSS-V

The last phase, SDSS-V, is currently taking data and focuses on "local" astrophysical observations. The Milky Way Mapper is a survey that aims to obtain optical and near-infrared spectra of more than four million stars located in the Milky Way. The Local Volume Mapper project will measure the spectra of galaxies in the local group, including the surrounding dwarf galaxies. Finally, Black Hole Mapper is a project that aims to obtain a better understanding of the nature of supermassive black holes inside quasars.

## 4.2.2 Instrument description

In this section, the telescope is described as it was used for the BOSS and eBOSS surveys. A picture and a schematic cut view of the SDSS telescope is given in Fig. 4.3. The telescope is of the Ritchey-Chrétien type with a primary mirror of 2.5 meters diameter. It has a field of view of 3 deg<sup>2</sup>.

### Spectroscopic mode

The spectroscopic mode is set up by placing a spectrograph fiber cartridge on the focal plane of the telescope, connected to two identical spectrographs as shown in Fig. 4.3 (right).

The spectrograph fiber cartridge is a system composed of a changeable pierced plate and manually inserted fibers. Aluminum plates on which fiber are placed are pre-drilled for each field of view according to the selected targets. A number of 1000 fibers connect the plate to the

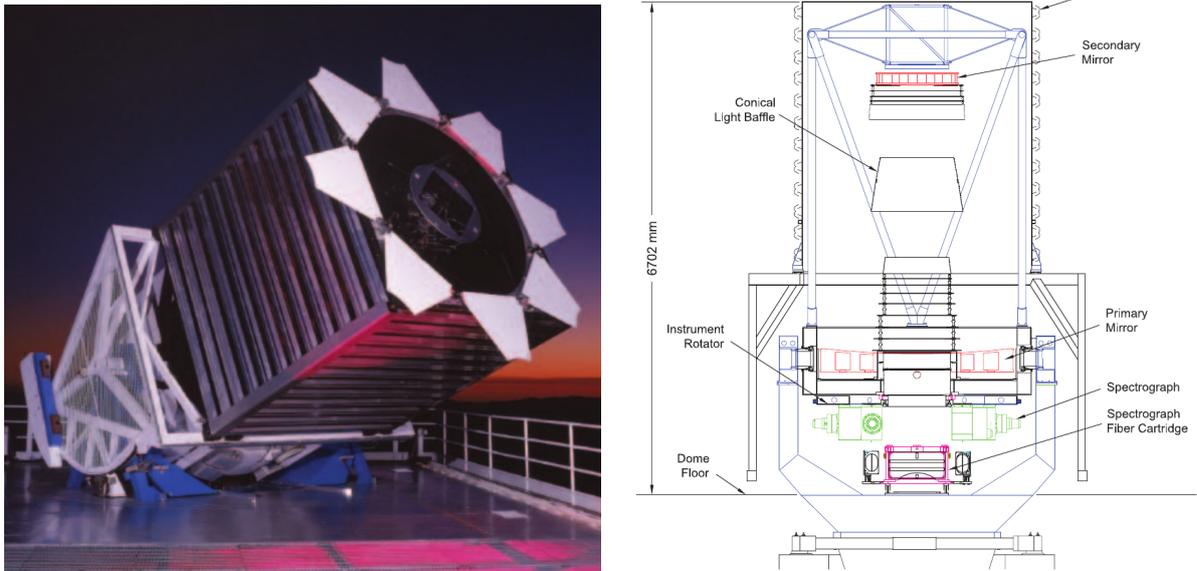


FIGURE 4.3 – (left) Photograph of the 2.5m SDSS telescope at night taken from [4] (right) Schematic representation of the different components of the SDSS telescope taken from [3]. On this drawing, the spectroscopic mode is set up.

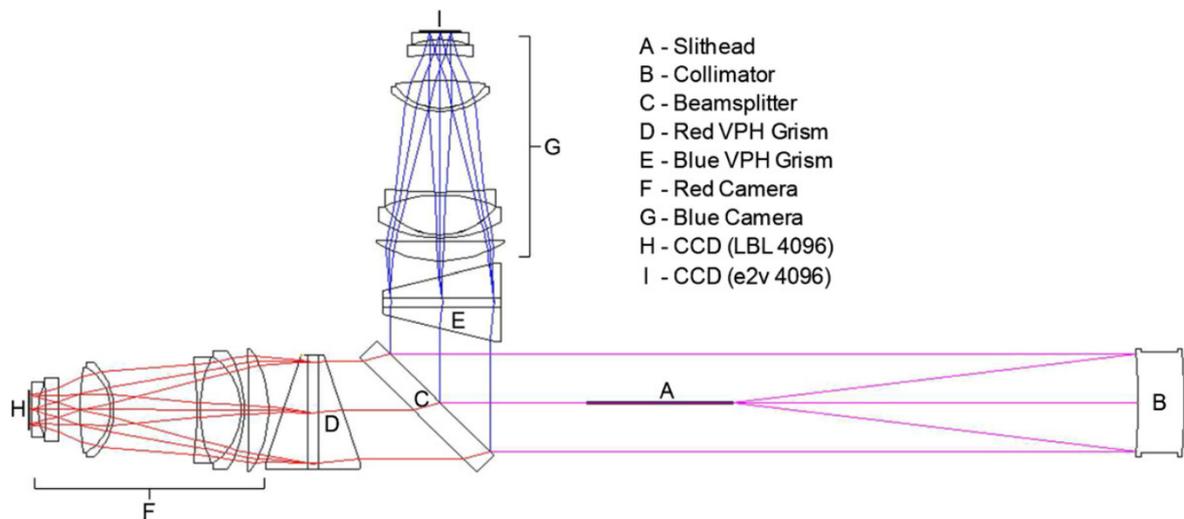


FIGURE 4.4 – Schematic of one of the spectrographs of the SDSS telescope taken from [3]

two spectrographs located on the side of the telescope. Each fiber has a diameter of  $120 \mu\text{m}$  which corresponds to an opening on the sky of  $2''$ .

A schematic representation of one of the spectrographs is given in Fig. 4.4. Each spectrograph receives light from 500 fibers through a slit plate mounted inside the slithead.

As shown in Fig. 4.4, the light coming from the fibers is transformed into a beam by a spherical collimator. A dichroic system composed of a beam splitter and two red and blue gratings split and direct the beam towards the red and blue cameras. These two cameras are each made of an optical system of 8 lenses and a  $4096 \times 4096$  pixels CCD. Each CCD is maintained to a temperature of  $180 \pm 0.5 \text{ K}$  by a cryostat.

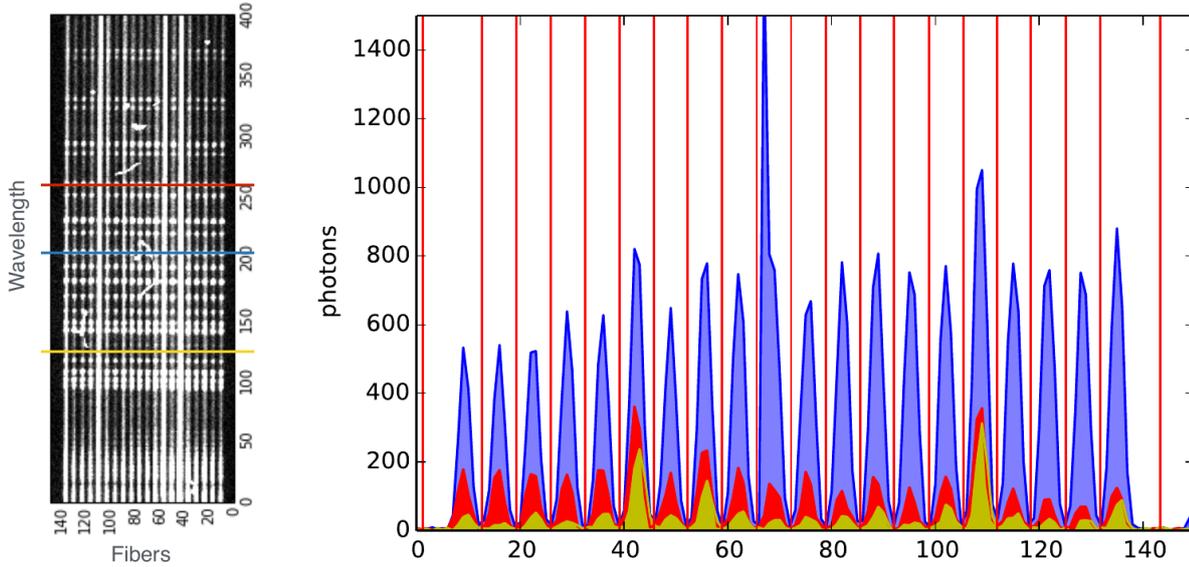


FIGURE 4.5 – (left) Example of raw image from the CCD of the SDSS spectrograph (right) One-dimensional cut of the flux from the raw image. The different colors match those of the image. Red lines represent the pixel separations when the boxcart extraction method is used. Source: Stephen Bailey, DESI internal presentation.

The blue camera covers the wavelengths range between 3600 and 6350 Å, and the red one between 5,650 and 10,000 Å. The lower wavelengths were chosen to allow the observation of the Ly $\alpha$  forest down to redshift  $z = 2.0$ . The higher redshift part allows to obtain information on the CaII, K and H absorption lines of galaxies up to redshift  $z = 0.8$ . These lines are useful for the identification of LRGs. The maximum and minimum values of the observable wavelengths are limited by physical considerations. The minimum value is limited by the UV atmospheric cut off around 3000 Å. The maximum wavelength is limited by the technology of the silicon-based sensors.

In this configuration the spectral resolution  $R$  of the spectrographs is between 1500 and 2300 for the blue cameras and between 1800 and 2700 for the red cameras.

### Photometric mode

To switch the telescope to photometric mode, the spectrograph fiber cartridge is replaced by an imaging camera made of an array of CCD with  $2048 \times 2048$  pixels. This array is composed of 6 columns of 5 CCD equipped with different filters  $r, i, u, z, g$  according to the lines. The central wavelengths of these filters are respectively 3590, 4810, 6230, 7640 and 9060 Å.

## 4.2.3 Spectroscopic pipeline

The images obtained from the spectrograph CCDs were processed with the `spec1d` software [15] written in IDL language. It is used to extract the spectra of the 500 fibres. The image from the CCD is first pre-processed with calibration images. The traces left by the cosmic rays are also suppressed.

Fig. 4.5 represents schematically the extraction of several pixels of the same row. In the right image, each color represents a row of the left figure. The spectrum is extracted from the image at each pixel using an optimal extraction method: for each pixel, a PSF-weighted profile is created.

The flux of a pixel is then given by the integral of the split profile. This method allows to take into account the cross-talk between fibers. However, it is a method implemented only along one direction and it is accurate only if the PSF profile can be decomposed along the two directions of the CCD. The flux is then calibrated using standard stars. The variance of the spectrum is calculated using the readout noise of the CCD and the number of incoming photons.

The spectra of the same object from different exposures are coadded and rescaled according to a logarithmic gridding with  $\Delta(\log(\lambda)) = 10^{-4}$ . The variance of the coadded spectra is obtained from the individual variances of each exposure. The typical exposure duration is approximately 15 minutes, and for eBOSS, the spectra of all Ly $\alpha$  QSOs were observed between 3 and 4 times each.

The `spec1d` software classifies the different spectra to identify the objects observed by eBOSS (star, nearby galaxies, LRG, ELG, QSO). The redshift as well as the type of the object is obtained by fitting a template decomposing in a sample base of known spectra. To assess the quality of the classification according to the different objects, the spectra were also visually inspected [16]. In particular, all QSOs observed in BOSS were visually inspected.

## 4.3 Dark Energy Spectroscopic Instrument

The **Dark Energy Spectroscopic Instrument (DESI)** is a project in the continuity of eBOSS. It had its first light in October 2019. Due to the Covid-19 pandemic, the telescope had a shutdown between March and October 2020. After a commissioning phase, the telescope observed continuously starting in December 2020, except for a planned shutdown in summer 2021. The DESI instrument is mounted in a previously used telescope called Mayall, now dedicated to DESI. It is located on the Kitt Peak National Observatory (KPNO) in the Tohono O'odham Nation <sup>1</sup> which is situated in the state of Arizona, US. The DESI survey is led by the Department of Energy (DOE) in the US office of Science, and in particular by the Lawrence Berkeley National Laboratory (LBNL). DESI is an international collaboration that includes more than 60 partner laboratories.

The DESI project is much more ambitious than its predecessor and relies on a change of all hardware and software previously used. Its objective is to observe the spectra of 35 million extra-galactic objects on a footprint of 14000 deg<sup>2</sup> in 5 years. This project aims to continue the cosmic mapping efforts made by SDSS, while drastically increasing its constraining power on the  $\Lambda$ CDM model and its possible extensions. The main goal of DESI is to constrain dark energy models with BAO. Additionally, constraints on neutrino mass and modified gravity will be achieved.

A major part of this thesis was dedicated to the understanding and analysis of the first DESI data. In particular, these data are used in Chap. 5 and 6. This section is based on the reference papers [17, 18], on internal DESI documents, on the previous thesis [19] and on my work. An overview of the 5 years of the DESI survey and its associated data will be presented in Sec. 4.3.1. Sec. 4.3.2 focuses on the description of the Mayal telescope and the DESI instrument. In Sec. 4.3.3, the course of a night of observation that I performed as a support observing scientist is reported. The DESI data processing pipeline is detailed in Sec. 4.3.4.

### 4.3.1 Survey overview

The DESI survey will last 5 years. The goal of DESI is to measure the spectra of about 35 million objects categorized in the same classes as eBOSS: bright galaxy sample (BGS), luminous red galaxies (LRGs), emission line galaxies (ELGs) and quasars (QSOs), as shown in Fig. 4.6 (left). The combination of all these targets will provide data over a very wide redshift range.

In parallel, some observation is focused on the observation of stars in our galaxy with the Milky Way survey (MWS). Also, some "secondary target" projects are carried out to maximize the scientific outcome of DESI, and also to prepare for future surveys. An example of a "secondary target" project I co-led is the observation of Lyman break galaxies (LBG) which are faint and located at high redshift. The goal is to demonstrate the feasibility of observing them in a potential "DESI-2" survey.

The scientific goal of DESI is to answer fundamental questions of cosmology, some examples of which are raised in Chap. 1: What is the nature of dark energy and dark matter? What is the composition of the Universe? How is the Universe structured?

These questions will be addressed first of all through the measurement of baryon acoustic oscillations (BAO) detailed in the Sec. 1.2.1.2 and the measurement of RSD explained in the

---

1. <http://www.tonation-nsn.gov>

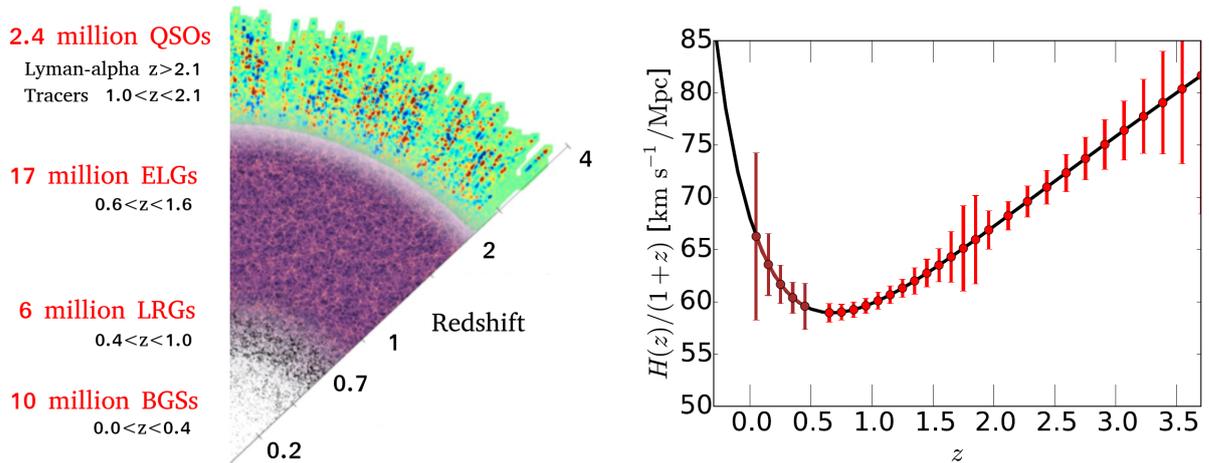


FIGURE 4.6 – (left) Expected number of spectra observed by DESI at the end of the survey. The classes of objects are targeted on different redshift ranges. Credits: DESI key figures for external communication. (right) Forecast on the DESI precision to measure the Hubble parameter using BAO as a function of redshift, taken from [17].

Sec. 1.3.3.1. In particular, the forecast on the ability of DESI to measure the Hubble parameter from the BAO distance measurements is shown in Fig. 4.6 (right). The measurement of a very large number of spectra will set constraints for a wide redshift range. DESI will notably improve the current growth constraints on  $f\sigma_8$  thanks to the observation of RSD as shown in Fig. 1.14.

The association of BAO and RSD measurements for an extensive redshift range will bring unprecedented constraints on the  $\Lambda$ CDM model. In addition, the large statistics of DESI will provide opportunities to constrain finer effects, such as modified gravity models, as shown in Fig. 1.14. DESI will significantly improve the current constraints on the sum of neutrino masses and potentially give a measurement. Finally, thanks to the large volume probed, DESI will be able to measure the potential non-gaussianity caused by inflation. This would put constraints on the primordial inflation phase of the Universe.

In parallel to DESI, many other surveys will also map the Universe on a very large footprint. The common observation areas with surveys like Euclid [20, 21], Rubin Observatory [1], or ACT [22] will bring unprecedented opportunities to perform joint analysis of different probes.

### Target selection

The first step of the DESI survey is the target selection (TS). This selection has been made beforehand and shall be frozen during the whole period of data collection of the main DESI survey. Unlike SDSS, the photometric data come from other telescopes. The photometric dataset used for the TS is the **DESI Legacy Imaging Surveys**. It is illustrated for the optical band r by Fig. 4.7. The data of this survey covers the 14,000 deg<sup>2</sup> of the DESI footprint and comes from the following surveys:

- **BASS/MzLS**: These two surveys image the north galactic cap (NGC) for declination above +34°. The Beijing-Arizona Sky Survey (BASS) started in 2015 on the 2.3m Bok telescope located at KPNO. It observed 5500 deg<sup>2</sup> with optical filters in the g and r bands. The Mayall z-band Legacy Survey (MzLS) used the pre-DESI configuration of the 4m Mayall telescope to observe the footprint covered by BASS in the z-band. This survey

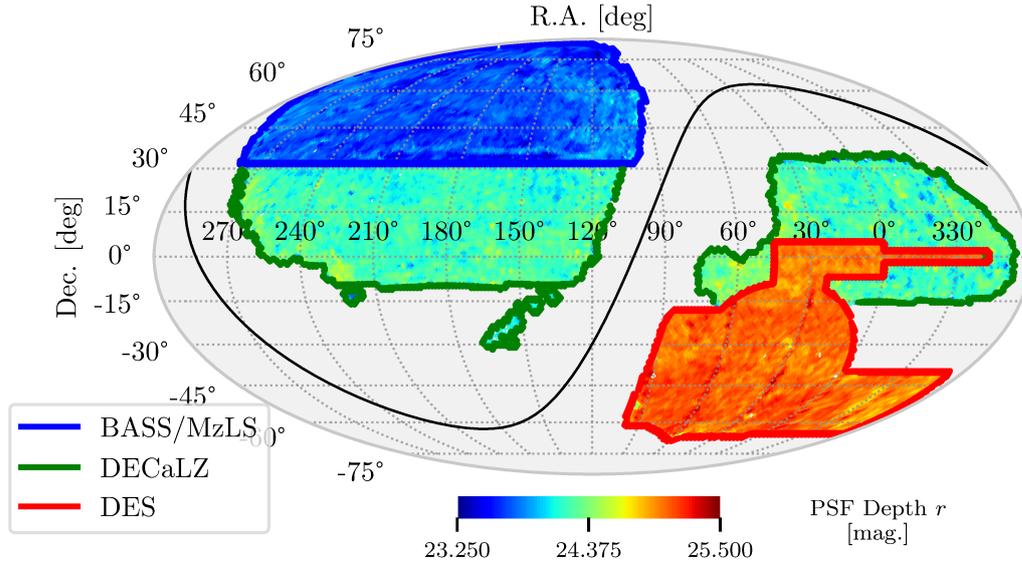


FIGURE 4.7 – Representation of the depth in the optical  $r$  band of the different surveys composing the DESI Legacy Imaging Surveys. The data forming this map come from the 9<sup>th</sup> data release of these surveys. The galactic plane is represented by a black solid line. This image was taken from [23]

started in 2016. Both surveys include an overlap area with DECaLS to study possible systematics differences.

- **DECaLS**: The south galactic cap (SGC) and part of the NGC was imaged by the Dark Energy Camera Legacy Survey. It started in 2014 and observed 4000 deg<sup>2</sup> with the three optical bands  $r$ ,  $g$  and  $z$ . This survey used the 4m Blanco telescope located at the Cerro Tololo Inter-American Observatory.
- **DES**: Dark Energy Survey (DES) data were used as external dataset. This survey of 5000 deg<sup>2</sup> in the SGC is designed to obtain weak lensing data with 8 optical bands. The three bands  $r$ ,  $g$  and  $z$  have been recovered for DESI imaging and the DECaLS survey has been designed not to re-observe the footprint covered by DES except on an overlap region used for systematics. The DES survey used the 4m Blanco telescope and started in 2013. The same camera was used for the DES and DECaLS surveys.
- **WISE**: The Wide-field Infrared Survey Explorer (WISE) satellite provided all-sky infrared observation over the bands W1 (3.4  $\mu\text{m}$ ) and W2 (4.6  $\mu\text{m}$ ).

From these data, the TS was performed in accordance with the desired number of observed objects for each class and their respective priorities. The TS uses a common pipeline detailed in [25]. Details of the selections are given for each class: BGS [26], LRG [27], ELG [28], QSO [23]. A more global overview of TS in the context of DESI scientific validation is given in [24].

The target selection for QSOs is mainly contaminated by stars, owing to their point-like shape. The use of infrared bands is essential for the QSO TS because these objects present a strong continuum in the infrared domain, not present for stellar spectra. The classical method for selecting quasars consists of making color cuts, i.e., cuts on the difference between two band magnitudes. The recent use of machine learning algorithms (random forest) improved the QSO target selection, especially at redshift  $z > 2.1$ .

Object class	Targets per deg <sup>2</sup>	Good redshifts per deg <sup>2</sup>
BGS	1366	971
LRG	615	548
ELG	2387	1144
QSO ( $z < 2.1$ )	208	140
Ly $\alpha$ QSO ( $z > 2.1$ )	100	60

TABLE 4.1 – Density of targeted objects and good redshift measured for each class of objects that DESI observes. The number of good redshifts was verified through the initial scientific verification phase of DESI. The numbers composing this table are taken from [24].

The statistics of target densities and good redshifts obtained from the DESI survey validation are reported in Tab. 4.1. For high-redshift QSOs ( $z > 2.1$ ), the achieved density is slightly larger than the forecasted one of 50 deg<sup>2</sup> [17]; thanks, in particular, to machine learning algorithms.

### Survey validation

The DESI collaboration conducted a survey validation (SV) from December 2020 to April 2021. A complete description of the SV is given in [24]. The objective of this phase was to test the performance of the instrument, the DESI pipeline and data quality, and to validate the survey scientific requirements related to the BAO and RSD measurements, as defined in [17]. To achieve this, observations different from those of the main survey were made.

In the validation survey, two main phases were realized. SV1 observations (from December 2020 to March 2021) include a large number of exposures for the same targets. This allowed among others, to study extensively the survey performance as a function of instrumental depth and to build visual inspection truth tables (see Sec. 4.3.4.3). SV3 phase (from April to June 2021) was dedicated more specifically to the evaluation of the survey design. The number of exposures is similar to the main survey and the goal was to determine the best strategy to cover the sky without fiber loss. The notations SV1 and SV3 are internal to the collaboration and do not appear in DESI reference articles. I use them to differentiate the types of data used in this thesis. A smaller set of observations called SV2 has been made but is not used for the analyses performed in this thesis.

During the SV, several reductions of the DESI pipeline were carried out. For each reduction, pipeline modifications were made and tested. The naming convention of the pipeline versions follow the names of famous mountains: Andes, Blanc, Cascades, Denaly, Everest, Fuji, and Guadalupe. The noise studies performed in this thesis have notably allowed to detect issues on the Blanc and Everest reductions. All the data collected during SV will be released in the early data release (EDR).

### Main survey

The DESI main survey started in May 2021. The period of data taking between this moment and July 2021, when the telescope was shutdown for hardware maintenance, is called data assembly 0.2 (DA0.2). Since September 2021, DESI has continued its observations without interruption. Data assemblies are planned to be made every year, and to be made public 6 months after their internal releases. The associated name is data release n (DRn) where n is the cumulative year of survey. In particular, DA0.2 data will be available in DR1.

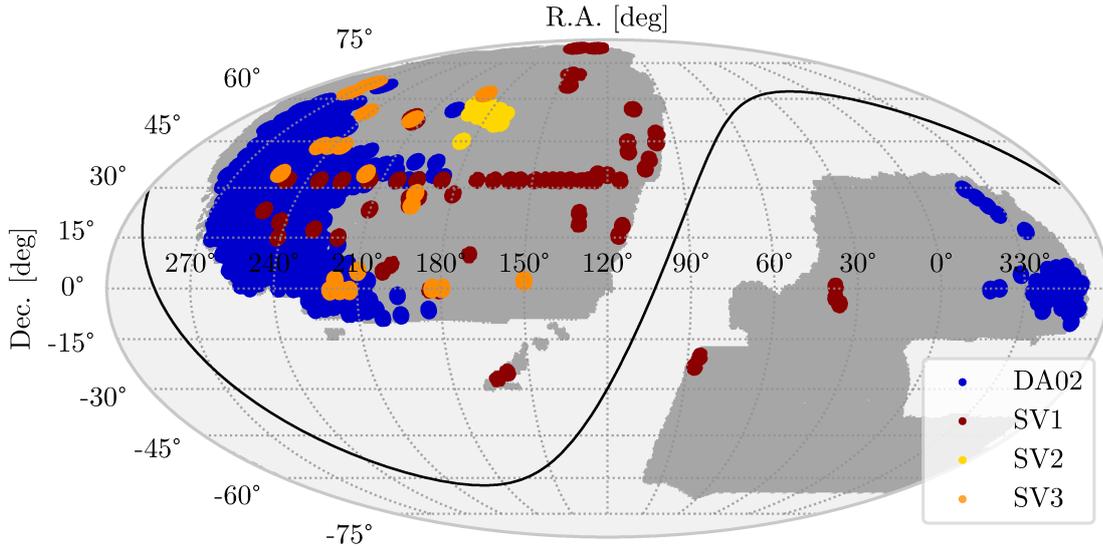


FIGURE 4.8 – Location on the sky of quasars whose spectra were measured during the SV and DA0.2 observations. SV2 is not used in this thesis. The black line represents the galactic plane, and the grey surface is the DESI final footprint. Credits: E. Chaussidon

In this thesis, the DESI data used are from SV1, SV3, and DA0.2. A representation of the quasar spectra measured during these observations is given in Fig. 4.8.

## 4.3.2 Instrument description

The DESI instrument is mounted on the Mayall telescope which is the largest one of KPNO. This observatory is supervised by NSF’s National Optical-IR Astronomy Research Laboratory (NOIRLab). A global illustration of the instrument and the telescope is shown in Fig. 4.9. The description that will be made in this section is mainly inspired from [18].

The global DESI system (instrument and telescope) is composed of:

- **Mayall telescope:** The global metal structure and the primary optical system
- **Prime focus corrector:** The optical corrector located at the top of the telescope.
- **Focal plane system (FPS):** A system positioned on the focal plane and which is composed of 5000 robotically controlled fibers.
- **Spectrographs:** A separated module containing 10 spectrographs.

### 4.3.2.1 Telescope

The detailed schematic design of the DESI instrument and its subsystems is provided in Fig. 4.10.

#### Telescope structure

The Mayall telescope is a reflective prime-focus telescope. As shown on Fig. 4.10, the light from the observed objects enters through the telescope dome and is reflected by a 4-meter diameter primary mirror. The beam is sent back to the top of the telescope after passing through the prime focus corrector. The beam is finally directed to the FPS.

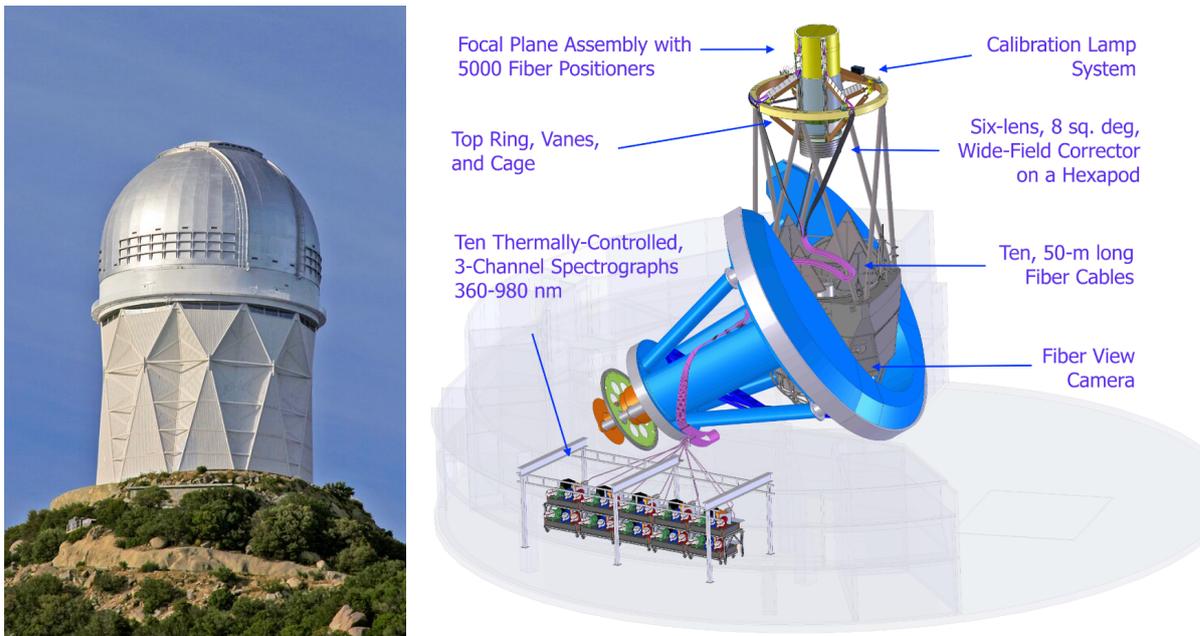


FIGURE 4.9 – (left) Photo of the Mayall telescope, the largest one of KPNO. Source: KPNO/NOIR-Lab/NSF/AURA/P. Marenfeld (right) 3D view of the DESI instrument mounted inside the Mayall telescope, taken from [29].

To maintain the FPS at the top of the telescope while providing a sufficiently large field of view, it was necessary to modify the Mayall telescope facilities. The details of these modifications are given in [30, 31]. A metal cage including the corrector and the FPS has been installed on top of the telescope. This cage is attached to the telescope by a large metallic ring.

### Prime focus corrector

The prime focus corrector is an optical system designed to increase the field of view on the focal plane. It is composed of a steel structure called corrector barrel and 6 lenses with diameters ranging from 0.8 to 1.14 meters [32]. For the range of wavelengths observed by DESI, its transmission is more than 90 %. Due to the use of a corrector, the focal plane surface has a small curvature which is taken into account during processing. The corrector focuses a total field of view of  $7.5 \text{ deg}^2$  on the FPS.

The corrector contains four spherical lenses, mounted in cells with flexure systems. The two last lenses constitute the atmospheric dispersion compensator (ADC). These are two wedged borosilicate lenses mounted in the barrel with one degree of freedom of rotation. They compensate the spectral spread of the incoming light that has passed through the atmosphere. Since this spread depends largely on the amount of atmosphere through which the light passes (air-mass), the ADC correction must be adapted by counter-rotating the two lenses according to the direction of observation.

When the telescope moves to different pointing directions, the presence of important masses and leverage arms generate consequent mechanical deformations. These deformations can lead in some cases to the misalignment of the corrector with respect to the telescope axis. To solve this issue, a hexapod connects the corrector barrel to the rest of the telescope. This system composed of six actuators is adaptive and maintains the alignment between the corrector and the primary mirror for every positions of the telescope.

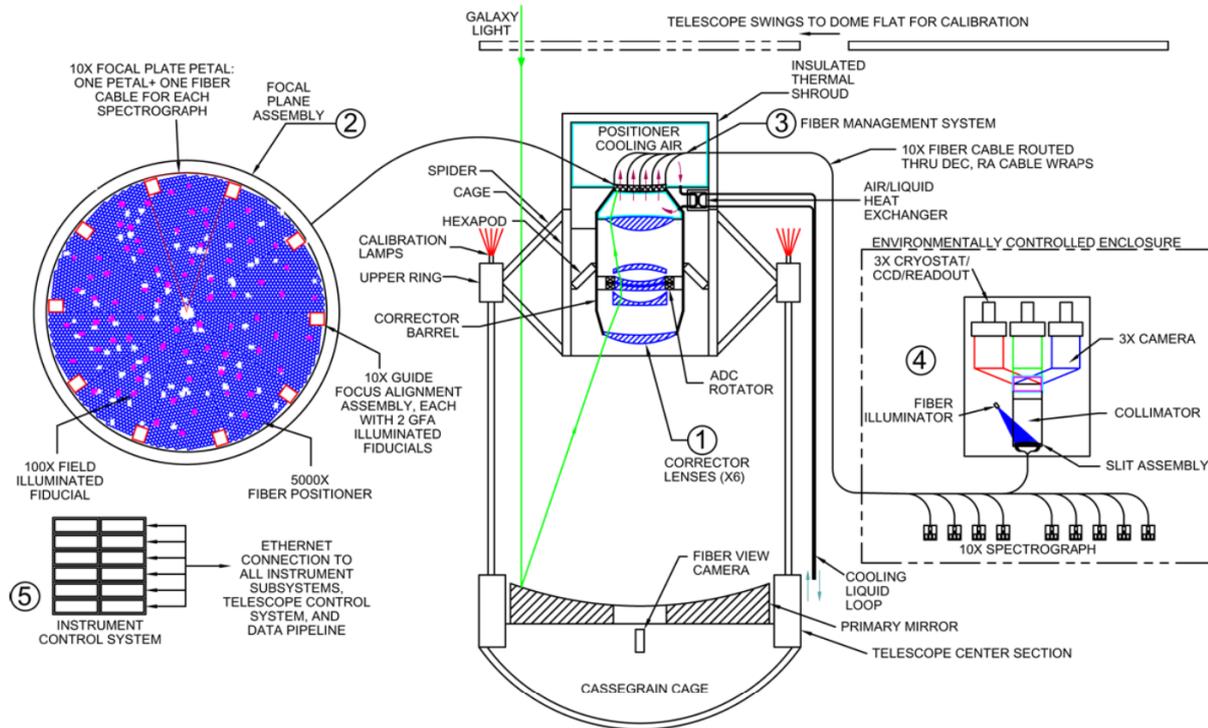


FIGURE 4.10 – Detailed diagram of DESI and the Mayall telescope taken from [18]

To install this system in the telescope, the corrector had to be aligned with great precision [33].

### Focal plane system

The focal plane system is the most important innovation of DESI. Compared to the SDSS telescope, where the optical fibers had to be plugged by hand, the system used for pointing the fibers to the sources is automated by positioners. An illustration of this system is given in Fig. 4.11.

The FPS is composed of three subsystems: the focal plane assembly, the focal plane enclosure and the fiber view camera (FVC).

The focal plane assembly is the system mainly illustrated in Fig. 4.11. It is composed of 10 identical modules called petals (shown in the right picture of the figure). Each petal contains 500 optical fibers attached to controllable robotic positioners. Each petal is equipped with a Guide, Focus and Alignment camera (GFA), between 10 and 12 field illuminated fiducials (FIFs) and two GFA illuminated fiducials (GIFs). In addition to all this, the assembly also includes the control electronics of these systems.

A positioner is a "theta-phi" robot whose tip can move with two rotational degrees of freedom. It allows to position a  $107\text{-}\mu\text{m}$  diameter fiber on a 3-mm radius disc surface. The set of all positioners allows to quickly modify the configuration of the FPS. While positioning 1000 fibers on SDSS could take one to two hours, positioning 5000 fibers on DESI can be done in a few minutes. The positioners are arranged in a hexagonal grid and separated by a 10.4 mm. Their positioning requires the use of an anti-colliding algorithm to prevent damage from contact between the robots.

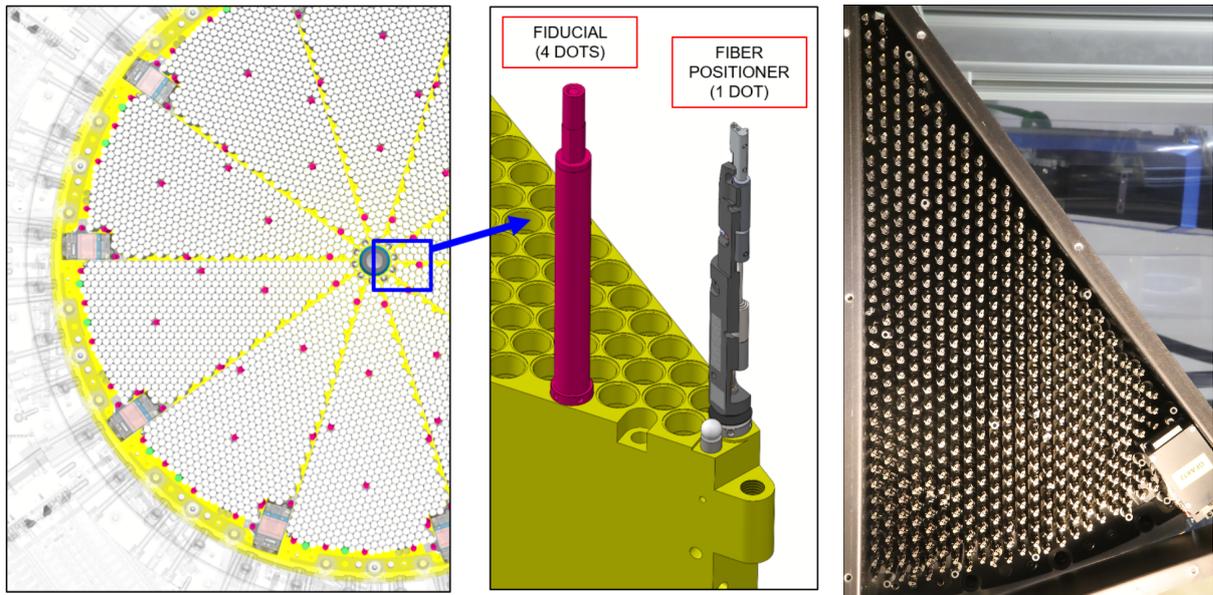


FIGURE 4.11 – (left) Global view of the focal plane system. The mechanical structure is shown in yellow. The fiducial fibers are represented in red and all the positioners in gray. GFA cameras located at the ends of each petal are shown in gray and pink. A 3D zoom of the FPS shows a fiber positioner and a fiducial fiber. (right) Photo of one petal. Both figure are taken from [34].

A positioning control loop ensures that all fibers are positioned at the right place. This requires the use of the FVC which is mounted in the middle of the primary mirror towards the FPS as shown in Fig. 4.10. This  $6132 \times 8176$  pixel CCD returns an image of the FPS with the distortions induced by the corrector. During a pointing, a first blind positioning is done without loop. Then, the fibers of each positioners are back-illuminated by LEDs. FIFs and GIFs are fibers that have a particular design and are illuminated by a specific light. They are used to identify respectively the FPS assembly and the GFA during the positioning control loop. The illumination of all the fibers gives a precise image of the position of each fiber and the deformations induced by the corrector thanks to the reference fibers. This image is used to correct the position of the fibers with a maximal rms error of  $5 \mu\text{m}$ .

The 10 GFAs of the FPS are cameras allowing the guidance, the focus and the alignment of the telescope. Six of them allow the guidance and the alignment by the observation of reference stars. The sphericity of their profile attests to the correct alignment of the FPS with respect to the primary mirror. The observation of these stars allows to slowly move the telescope to follow the sky. Four other cameras are used for the focus. For this, a dual thickness optical filter place the stars about 1.5 mm above and below the focus of the camera. The observed distorted stars appear as donuts on the CCD. A spherical and undistorted profile attests the good focus of the instrument.

The focal plane enclosure takes care of protecting the assembly and keeping the temperature of the positioners at reasonable values. On the one hand, overheating can damage the positioners, and on the other hand, significant temperature fluctuations can distort the image.

From each petal, the 500 optical fibers are grouped in a fiber cable which is connected to a spectrograph. Each spectrograph is associated with a petal of the FPS. The fiber cables have a total length of about 47 meters. The path of these cables along the telescope was designed

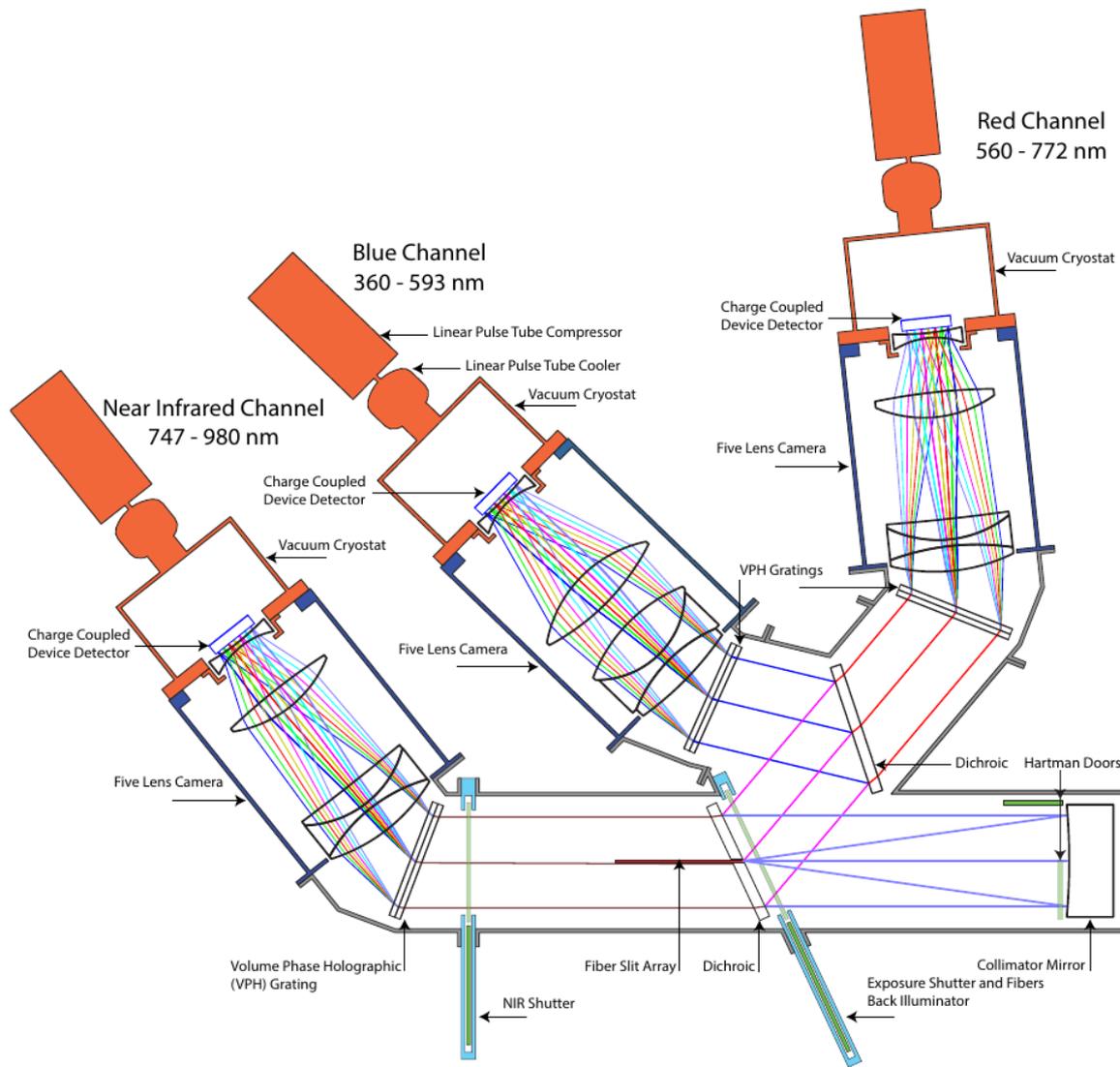


FIGURE 4.12 – Schematic representation of one spectrograph of the DESI instrument taken from [18]

in order to avoid imposing a too large bending radius on the fibers and to keep the telescope's freedom of movement.

### 4.3.2.2 Spectrographs

The cable containing 500 fibers is brought to the fiber split of one of the spectrographs described in detail in [35]. The fibers are aligned and their beam is sent to the optical system of the spectrograph. This system is schematically represented in Fig. 4.12.

Just after the fiber slit, an exposure shutter is installed to prevent light from entering the optical system. This shutter is also equipped with LEDs. These are used to back-illuminate the optical fibers during the positioning control loop. When the shutter is open, the light from fibers is collimated by a mirror of focal ratio  $f/3.57$  in order to obtain a parallel beam. The beam is sent to a dichroic separating the light between the near infrared (NIR or Z) camera and the others, while a second dichroic separates the beam between the red (R) and blue (B) cameras.

Channel	Spectral range (Å)	Spectral resolution
Blue (B)	3600 - 5930	2,000 – 3,200
Red (R)	5600 - 7720	3,200 – 4,100
Near Infrared (Z)	7470 - 9800	4,100 – 5,100

TABLE 4.2 – Spectral range and resolution for each channel of the spectrographs taken from [18]. In comparison to SDSS spectrographs, the resolution improved by almost a factor of two.

The dispersion of light from the fibers for its spectral analysis is realized by a volume phase holographic grating. Each of these gratings is made by holographic printing. They have a large grating area and deliver a high transmission (85 to 90 %) over a wide wavelength band. The dispersed light passes through a refractive optic system of 5 spherical lenses (3 less than SDSS camera) with two aspherical surfaces.

Finally, the light is captured by CCD sensors of  $4096 \times 4096$  pixels. The spectral resolution of the instrument is driven by the point spread function (PSF), which is the response of an optical system to a point source, of these cameras taking into account the optical system. This resolution was measured beforehand by the image given by a light source of  $12 \mu\text{m}$  diameter. The result of the resolution measurements is given in Tab. 4.2. The constraints on the total spectral range are the same as for the SDSS camera, for observing the Ly $\alpha$  forest at  $z = 2.0$  and the LRGs. The resolution was improved for all cameras. The main motivation of the near-infrared camera improvement is to discern the OII doublet of ELGs. The resolution improvement for the blue camera by almost a factor of two compared to eBOSS is very important in this thesis. Indeed, for the measurement presented in Chap. 5, the resolution drives the ability to probe the smallest scales of the intergalactic medium.

Finally, cryostats maintain the vacuum and a low temperature in the CCD chamber to minimize the read-out noise of cameras. These cryostats are set to ensure a fixed temperature of 163 K for the blue CCDs and 140 K for the other two types of CCDs. The vacuum is maintained in the chamber at  $10^{-6}$  mbar.

### 4.3.3 Observing with DESI

During my thesis, I had the opportunity to be a support observing scientist of DESI for 4 half nights. This subsection describes the course of an observing night, emphasizing the use of the different systems described in previous section.

During its nominal operation, the telescope is never completely shut down. For example, the cryostats always maintain the vacuum and the temperature of the CCD to avoid condensation issues. The first part of a night of observation consists in checking the good functioning of all systems in operation. A verification of the correct functioning of the various cameras (spectrographs CCD, GFA, FVC) is carried out.

An important stage of calibration of the spectrograph cameras is then carried out. For these calibrations, the telescope’s dome is closed, and there is no additional light inside. Calibrations are performed each night to subtract the background noise present in the CCD’s signal. When the weather prevents observation, detailed monthly calibrations are done with varying exposure time. In each case, the exposures performed are the following:

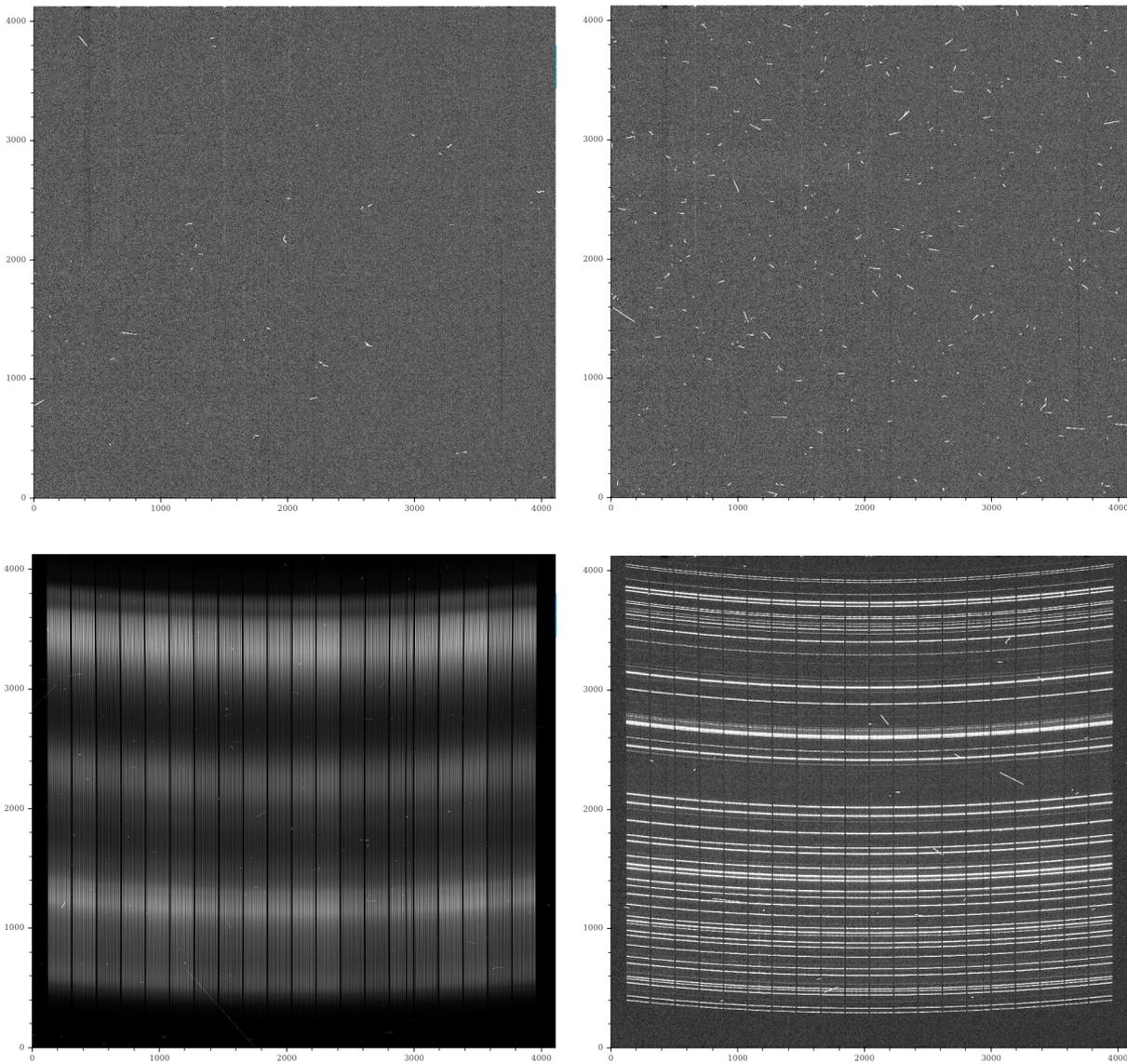


FIGURE 4.13 – Calibration images during one of my nights of observation (January 12, 2022) for the red camera of the DESI spectrograph number 0. The axes show the pixel position in the camera. The different fibers are arranged along the x direction, and the wavelength dispersion of each fiber along the y direction. (left up) Zero/bias frame *preproc-r0-00118258.fits* (right up) dark frame *preproc-r0-00118262.fits* (left bottom) flat frame *preproc-r0-00118299.fits* (right bottom) arc frame *preproc-r0-00118271.fits*.

- **Zero/bias:** A CCD has an intrinsic readout noise. This noise depends on the position on the CCD. To subtract it, an image with the minimum response time of the CCD is taken and then subtracted.
- **Dark:** During an observation, the CCD produces electronic noise called dark current. An image with the exposure shutter covering the fiber slit is taken during approximately the same exposure time as the observations. The subtraction of this image during processing suppresses this associated noise.
- **Flat:** A screen inside the dome is illuminated in a uniform way by a calibration lamp with a smooth spectrum. The telescope is then pointed towards this screen. Each pixel of the spectrograph CCD receives the same light flux at a given wavelength. This flat

exposure allows to correct the relative response of each pixel to an incoming light flux. During processing, the flats are used to normalize the received flux over the whole CCD.

- **Arc**: A last step is the observation by the telescope of clearly defined emission lines. This is achieved by projecting light from calibration lamps onto the dome screen. The image obtained called arc is used during processing for wavelength calibration and to characterize the spectral resolution as detailed in Sec. 4.3.4.1.

For flats and arcs, the telescope is pointed towards a white point inside the dome on which the calibration lights are projected. An example from one of my observation nights of the 4 calibration image types is given in Fig. 4.13. To remove statistical fluctuations, all calibrations are repeated several times and their averages, usually called **masters**, are used to process the CCD images.

After obtaining CCD calibration images, the observation starts at about 12° twilight, which means that the Sun is 12° below the horizon. The end of the observations is usually done with the same angular position of the Sun on the other side.

To observe, the telescope is driven by a software called instrument control system (ICS). It manages DESI subsystems and also collects the signals from all the sensors. This system takes care of telescope movements. All the security measures to ensure the quality of the data and to maintain the integrity of the telescope are also managed by the ICS.

A specific sky field with the same size as DESI field of view is called a **tile**. At a given time of the night, the choice of which tile is observed depends a lot on the atmospheric conditions estimated by telemetry. The ICS records conditions such as the sky brightness, seeing, transparency and throughput. A software called effective time calculator (ETC) determines an effective observation time based on these measurements. The ratio between the effective exposure time and the actual observation time is called survey speed. A software called next tile selector (NTS) selects the next field of view and target type to observe on the basis of the integrated survey speed over 20 minutes and the progress of the DESI survey. Depending also on the moon position, the most favorable atmospheric conditions are used to observe faint sources (dark program). When the sky is less clear, brighter targets are observed (bright program).

To observe a tile, the telescope is pointed in its direction, then the corrector is aligned and focused using the images of bright stars on the GFA. The FPS fibers are precisely positioned using the positioning control loop. During an exposure that lasts about 10 minutes, the GFA are used to guide the telescope. Active guidance is needed to correct for the Earth's rotation, but also for the wind that can deviate the telescope.

Finally, at the end of an exposure, we obtain a raw CCD image as shown in Fig. 4.14 (left). This image represents the number of photons captured at each pixel of the CCD. It is obtained after subtracting the over-scan and pre-scan areas of the CCD. The different fibers are distributed on the x-axis while the spectra for different wavelengths are dispersed along the y-axis. On this image, the spectra of 500 fibers grouped by packages of 25 are displayed. In this picture, similarly to the calibration arc image, horizontal lines coherent between fibers appear. These features come from emission lines of the atmosphere called **sky lines**. We also see parasite cosmic rays hitting the CCD in different random directions. Cosmic muons cause the dots and straight lines. Curved lines, otherwise called worms, are produced through Compton-scattering of gamma rays or beta emitters near the telescope.

Raw images are corrected using calibration images, with the DESI pipeline. My role during my nights of observation was to assess the quality of the exposures by checking the estimated

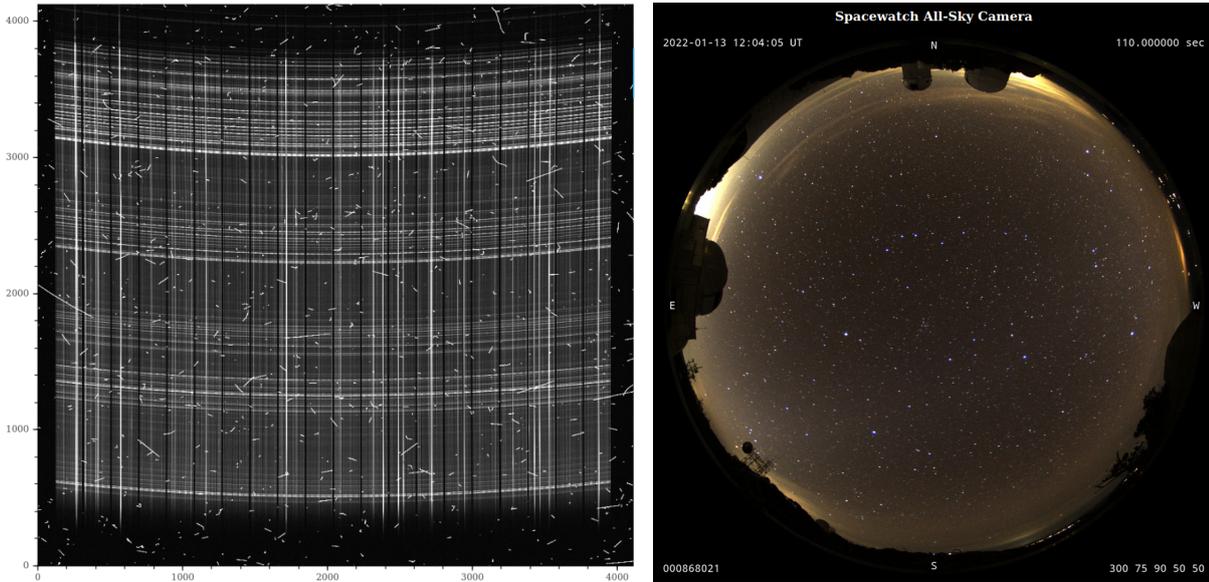


FIGURE 4.14 – (left) Raw CCD frame *preproc-r0-00118349.fits* during one of my nights of observation (January 12, 2022) for the red camera of the DESI spectrograph number 0. The axes show the pixel position in the camera. (right) All-sky camera image taken during the same exposure. This camera is used to have a global view of the sky at Kitt Peak. Polluting light at the edge of the camera comes from cloud reflection of Phoenix and Tucson public lighting. The image contrast is very high compared to eyesight and the polluting light are not as bright as they appear on this image. At this moment, the atmospheric conditions were excellent and the survey speed was 120%.

noise on the CCDs, the good guidance of the telescope, the atmospheric conditions and a quick estimate of the extracted spectra detailed in the next section.

## 4.3.4 DESI spectroscopic pipeline

### 4.3.4.1 Spectroscopic extraction

Spectroscopic extraction refers to the set of methods that aim to obtain several spectra such as in Fig. 2.3 from a CCD image as in Fig. 4.14. For DESI, the software used for the processing of the CCD images are referenced in the `desihub`<sup>2</sup> repository. The detail of the pipeline is reported in [36].

#### Pre-processing

Before extracting the spectra, raw images are corrected using the calibration images. The dark and bias images are subtracted to remove associated readout noises. The non-uniform CCD pixel response is corrected with the flat image. This procedure is called pixel flat fielding. Finally, the CCD over-scan readout noise is removed. The physical origins of the CCD noise sources is detailed in Sec. 4.3.4.2.

In a second step, a software is used to detect and remove defective or saturated pixels, and traces left by cosmic rays.

#### Algorithms for spectroscopic extraction

2. <https://github.com/desihub>

From a cleaned CCD image, it is possible to extract the spectra of each fiber using several techniques. For DESI, two algorithms are implemented to realize this extraction. An online monitoring called `nightwatch` is used to quickly estimate spectra during the observation nights for quality assessment. A more complex and time-consuming algorithm is used to obtain the spectra for the main scientific analyses. It is based on the "spectroperfectionism" methodology [37], and implemented in the `desispec`<sup>3</sup> package.

The `nightwatch` software uses a boxcart extraction. The separation between two spectral lines is defined where the flux is minimal. The flux value of one pixel is then the sum of the photons in each separated bin. This is a robust and simple technique to obtain spectra within minutes after the data collection, and do quality assessment very quickly. However, this method has some disadvantages. First, when two pixels overlap, the cross-talk between these pixels is not correctly taken into account in the spectroscopic extraction. This miscalculation is even more important when a highly illuminated pixel is next to a poorly illuminated one. Furthermore, the gridding in wavelengths is entirely defined by the columns of the CCD and not linked to the PSF of the observation. This creates unwanted differences between the fibers.

The "spectroperfectionism" method is an improvement of the optimal extraction algorithm of the SDSS telescope detailed in the section 4.2.3. The main problem with this extraction lies in its one-dimensional nature. Indeed, if the PSF profile depends simultaneously on both directions of the CCD, the SDSS extraction method is inaccurate. This problem is circumvented in the DESI pipeline by performing a 2D PSF spectroscopic extraction. The detail of this extraction is given in [37].

In the "spectroperfectionism" paradigm, the raw camera signal can be mathematically modeled by

$$\vec{p} = \mathbf{A}\vec{f} + \vec{n}. \quad (4.1)$$

Starting here, we choose the convention to represent matrices in bold when only 3-dimension vectors are considered. The CCD pixel count  $\vec{p}$  is a one-dimensional vector containing the signal from all the pixels. It is decomposed as the noiseless signal  $\mathbf{A}\vec{f}$  and a pixel noise vector  $\vec{n}$ . In the noiseless signal, the matrix  $\mathbf{A}$  is very sparse and contains all the information concerning the shape of the PSF and the wavelength calibration.  $\vec{f}$  is the input spectrum vector, i.e. the signal that we want to recover.

The matrix  $\mathbf{A}$  can be measured thanks to the arc images presented previously. The arc emission lines allow to measure the shape of the PSF and its dependence on the position and the wavelength. The shape of the PSF is obtained by fitting a linear decomposition in the Gauss-Hermite basis:

$$\text{PSF}(x, y) = \sum_{i,j} a_{i,j} H_i \left( \frac{x - x_c}{\sigma_x} \right) H_j \left( \frac{y - y_c}{\sigma_y} \right) \times \frac{1}{2\pi\sigma_x\sigma_y} \exp \left( -\frac{(x - x_c)^2}{2\sigma_x^2} - \frac{(y - y_c)^2}{2\sigma_y^2} \right), \quad (4.2)$$

where  $H_i$  is a  $i$ -degree Hermite polynomial,  $x_c$  and  $y_c$  are spectral trace coordinates, and  $\sigma_x$  and  $\sigma_y$  are the Gaussian width along both directions. It is thus in the  $\mathbf{A}$  matrix that the resolution properties are encoded.

The principle of spectroscopic extraction is to obtain  $\vec{f}$  from  $\vec{p}$  knowing  $\mathbf{A}$ . It is possible to do this by assuming the noise as Gaussian. We then obtain the following minimal- $\chi^2$  solution:

---

3. <https://github.com/desihub/desispec>

$$\vec{f} = (\mathbf{A}^t \mathbf{N}^{-1} \mathbf{A})^{-1} \mathbf{A}^t \mathbf{N}^{-1} \vec{p}, \quad (4.3)$$

where  $N_{ij} = \langle n_i n_j \rangle$  is a diagonal matrix because the raw pixel errors are treated as statistically independent. This step defines a limit on the sampling for  $\vec{f}$  with respect to the wavelengths. Indeed,  $\mathbf{A}^t \mathbf{N}^{-1} \mathbf{A}$  must remain invertible and a too fine sampling prevents that. The inverse matrix  $\mathbf{C} = (\mathbf{A}^t \mathbf{N}^{-1} \mathbf{A})^{-1}$  also turns out to be the covariance matrix of the estimator  $\vec{f}$ . At that stage the gridding in wavelength is chosen for all DESI observations to be 0.8 Å.

The issue which persists after this operation is that Eqn. 4.3 de-convolves the spectral resolution, leading to unstable numerical features in the computed vector  $\vec{f}$ . The solution used is to re-convolve  $\vec{f}$  to the same resolution as the raw data.

As  $\mathbf{C}^{-1}$  is symmetric, positive definite and band-diagonal, we can define its square root  $\mathbf{Q}$  such that  $\mathbf{C}^{-1} = \mathbf{Q}\mathbf{Q}$ . This matrix is used to diagonalize  $\mathbf{C}$  and the diagonal matrix  $\tilde{\mathbf{C}}$  is obtained by

$$\tilde{\mathbf{C}} = \mathbf{R}\mathbf{C}\mathbf{R}^t, \quad (4.4)$$

where  $\mathbf{R}$  is the **resolution matrix**. It contains all the information necessary to characterize the resolution of DESI spectrographs. It is defined by

$$R_{ij} = \left( \sum_j Q_{ij} \right)^{-1} Q_{ij}. \quad (4.5)$$

Here, the Einstein summation convention is not used. It is then possible to define the re-convolved spectrum vector by

$$\vec{f}_c = \mathbf{R}\vec{f}. \quad (4.6)$$

This spectrum vector corresponds to the spectrum that would have been observed by a one-dimensional spectrograph with the same resolution as the actual two-dimensional spectrograph of DESI. It is the final product of the spectroscopic extraction. The error model composed of the uncorrelated errors is then given by the diagonal matrix  $\tilde{\mathbf{C}}$ . Consequently, the one-dimensional  $\vec{f}_c$  pixels are statistically independent of other spectral lines. For simplicity, this measured flux will be written  $f$  hereafter.

This method has many advantages in addition to correctly modeling a complex PSF. It also provides the encoding for each fiber and each wavelength of the non-Gaussian instrument resolution thanks to the matrix  $\mathbf{R}$ . With this method, all the spectra are defined on the same wavelength grid without additional resampling. All the information of the CCD is preserved. Finally, the output stream is uncorrelated between wavelength bins. The downside of all these advantages is that the method is very expensive numerically, and that any bad pixel on a CCD may impact several spectra and wavelengths.

### Spectral calibration

Once the spectrum of each fiber is extracted, several post-treatments allow to remove observational effects. For all exposures, some fibers do not point to any target and are dedicated to the observation of the sky. The spectra associated with these sky fibers provide an estimate of the intensity of the sky lines that are subtracted from the spectra associated with the targets.

Similar to CCD pixels, the response of individual fibers is not uniform. The flat frames determine how the input flux depends on the fiber. Normalizing the flat image, called fiber flat fielding, corrects the fiber's non-uniformity.

The transmission defaults of the atmosphere and of the telescope as a whole are corrected by the observation of calibration stars. The spectrum of these main sequence stars is known and modeled from previous observations. The transmission is therefore corrected using the ratio of their DESI spectra. This step provides an absolute calibration scale for the spectra. It is used to transform CCD units (number of electrons) to observed flux units ( $10^{-17} \cdot \text{erg} \cdot \text{s}^{-1} \cdot \text{cm}^{-2} \cdot \text{\AA}^{-1}$ ).

### Exposure coadding

After the flux calibration step, cleaned spectra are obtained for each exposure. Many objects are observed during several exposures. Their spectra must therefore be coadded. We consider an object with a number  $N_{\text{exp}}$  of individual exposures. We note  $f_k$  the measured flux of exposure  $k$ . The coadded spectrum is defined by

$$f = \frac{\sum_k (\mathbb{V}_{\text{pip},k})^{-1} f_k}{\sum_k (\mathbb{V}_{\text{pip},k})^{-1}}, \quad (4.7)$$

where  $\mathbb{V}_{\text{pip},k}$  is the variance of  $f_k$  detailed in Sec. 4.3.4.2. The standard deviation of the coadded spectrum is given by

$$\sigma_f = \frac{1}{\sqrt{\sum_k (\mathbb{V}_{\text{pip},k})^{-1}}}. \quad (4.8)$$

During coaddition, masked pixels affected by CCD defects or cosmic rays must be taken into account.

#### 4.3.4.2 Evaluating the noise

##### Pipeline noise

As explained previously, noise is mathematically introduced into the DESI pipeline by adding the pixel noise vector  $\vec{n}$ . It results from the addition of effects modeled at the CCD scale. Each contribution is calculated by measuring the spatial variance on the CCD image from the associated noise source. It is assumed that this noise comes from the following four sources:

- **Poisson noise:** Measuring photons with a CCD is a statistical process. It creates a noise source which is directly linked to the input flux, and particularly dominant for low fluxes. For DESI, this noise is estimated by modeling the CCD. All the other noise sources are flux-independent.
- **Over-scan:** The over-scan is the covered part of the CCD chip used to measure the bulk offset, i.e., the average level of all pixels. It is modified by temperature and is used to remove small variations in the bias. Over-scan suppression introduces noise.
- **Bias:** Noise due to the response of the CCD to a minimal exposition time. It is associated with the bulk offset's pixel variation and emerges from parasite electron or pixel defect. The master bias is used to estimate this noise.
- **Dark current:** Readout noise due to the thermal motion of the atoms composing CCD material. This motion induces charge deposit on the CCD. Dark current is estimated using the master dark.

By adding these four terms, we obtain a CCD noise estimator noted  $\vec{n}$ . It is then propagated to the 1D spectra by the "spectroperfectionism" formalism to a **pipeline noise** vector  $\sigma_{\text{pip}}$ , with a size given by the number of wavelengths  $N_\lambda$ . We will note the variance vector  $\mathbb{V}_{\text{pip}}$ . In the DESI pipeline, this vector is stored as the inverse variance noted IVAR or  $\mathbb{V}_{\text{pip}}^{-1}$ .

This pipeline noise estimation is based on direct measurements of known effects in the CCDs. As a drawback, some other contributions to the noise may be omitted, so that the noise would be underestimated.

### Test of the pipeline

It is possible to characterize the noise level independently from the pipeline through a statistical evaluation. The accurate estimation of the noise level is of particular importance for the measurement of the one-dimensional power spectrum of the Ly $\alpha$  forest detailed in Chap. 5. Furthermore, since DESI is a new instrument, it is important to first validate the noise modeling. For these reasons, I developed a statistical method during my thesis to cross-check the noise level estimated by the pipeline.

This method consists in using multiple exposures of luminous targets. During my thesis, I used QSO and LRG spectra that had many different exposures in the first DESI validation data. I created a noise estimator called **diff noise** using the difference between the individual exposures. The diff noise is computed for objects with an even number  $N_{\text{exp}}$  of individual exposures. Each exposure has an associated pipeline variance  $\mathbb{V}_{\text{pip},k}$ . The flux difference of this object indexed by  $j$  is given by

$$\Delta f_j = \frac{\sum_{k=1}^{N_{\text{exp}}} (\mathbb{V}_{\text{pip},k})^{-1} (-1)^k f_k}{\sum_{k=1}^{N_{\text{exp}}} (\mathbb{V}_{\text{pip},k})^{-1}}. \quad (4.9)$$

We define the average diff noise  $\overline{\mathbb{V}}_{\text{diff}}$ , for an observed number of objects  $N_{\text{obj}}$ , by the variance of the  $\Delta f_j$  differences:

$$\overline{\mathbb{V}}_{\text{diff}} = \frac{1}{N_{\text{obj}}} \sum_{j=1}^{N_{\text{obj}}} (\Delta f_j - \overline{\Delta f})^2, \quad (4.10)$$

where  $\overline{\Delta f}$  is the average of the  $\Delta f_j$  for all objects.

The average diff noise  $\overline{\mathbb{V}}_{\text{diff}}$  can be compared to the average pipeline noise  $\overline{\mathbb{V}}_{\text{pip}}$  for the same objects and exposures. In order to be less sensitive to outliers, one may compute these averages using percentiles instead of the variance in Eqn. 4.10 and compare it to the median value of the pipeline noise. An example of comparison between these two outlier-insensitive estimators is given in Fig. 4.15.

It is important to note that the normalization by the inverse variance in the calculation of the exposure difference given by Eqn. 4.9 may introduce an error if the pipeline noise estimator is misestimated by an additive correction. However, this effect is secondary and tends to be reduced when the pipeline noise estimate is improved.

Using the diff estimation method, I found that the pipeline noise was underestimated in the first reductions of the DESI validation data. At this stage, we have an estimation of this under-estimation for all the wavelengths. The DESI pipeline team looked for ways to improve this noise estimate. It was found that there was a source of noise that had not been previously taken into account. This missing term is the Poisson noise from the dark current that is part of a zero exposure (source: J. Guy). Secondly, the CCD position dependence of the noise improved.

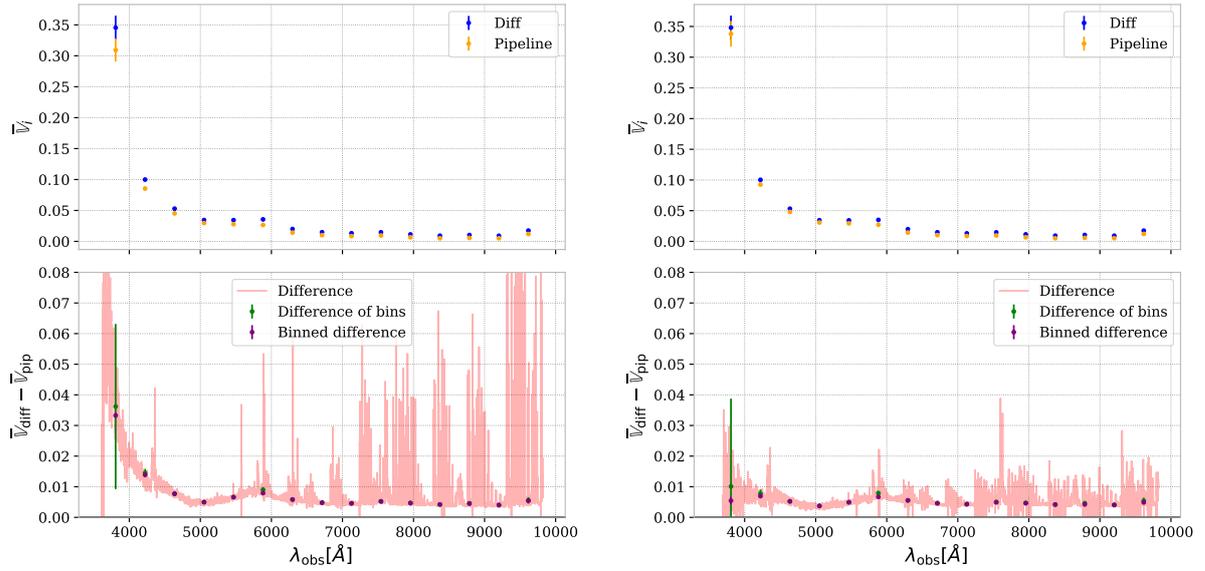


FIGURE 4.15 – Comparison between the average diff and pipeline noise. The top panels show the average variance for both estimators. The bottom panel picture the estimator difference with and without a wavelength binning. The panels on the left are run with data from the Everest pipeline reduction and on the right with the more recent Fuji reduction. Noise estimators are calculated using LRG and QSO spectra. Fourteen SV1 tiles with large exposure time are selected (80613, 80702, 80674, 80605, 80686, 80690, 80684, 80609, 80694, 80610, 80608, 80678, 80607, 80606), and the spectra from all spectrographs are used.

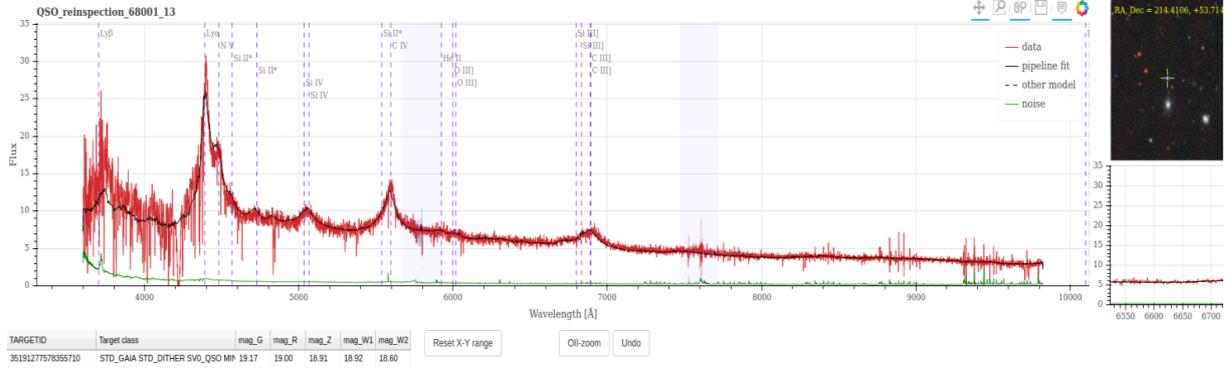


FIGURE 4.16 – Screenshot of the visual inspection of a QSO spectrum at  $z = 2.61$  from the first DESI data, shown thanks to the `prospect` software. The flux is shown in red, the `redrock` best-fit in black and the pipeline noise in green.

The left and right panels of Fig. 4.15 illustrate the improvement brought by the new noise estimate: in the latest data reduction (Fuji), the agreement between  $\bar{V}_{\text{pip}}$  and  $\bar{V}_{\text{diff}}$  is clearly improved, especially in the blue part of the spectrographs. There is still a difference however, at the 10 % level.

### 4.3.4.3 Classifying spectra and measuring redshifts

The classification of observed spectra in different categories, as well as the determination of their redshifts is realized by the `redrock` software included in the `desihub`<sup>2</sup> repository. The complete description of this software is given in [38].

The `redrock` software is similar to the one used by SDSS detailed in the Sec. 4.2.3. For each class of object (Star, galaxy and QSO), templates are built from previous observations. Compared to the software used by SDSS, the shape of these templates is built from the SDSS spectra which gives a better variety of spectral types. A principal component decomposition allows to obtain a linear basis of these templates. For each observed spectrum, a linear combination of this basis with variable weights is fitted, and this for several redshifts. By selecting the fit with the lowest  $\chi^2$  between this linear decomposition and the spectrum, the class and the redshift of the spectrum are simultaneously determined. This is an expensive algorithm because of the realization of fits at many different redshifts on a lot of spectra. It is therefore usually run in multi-processing. In the particular case of QSOs, additional algorithms are used to improve the completeness of the catalog. This will be described in Chap. 5.

In the early phases of DESI, to assess the quality of the spectral classification and redshift determination, a **visual inspection (VI)** step was performed. During my thesis, I participated to this step for the QSO targets. The `prospect` software in the `desihub`<sup>2</sup> repository allows to read the spectra together with the files generated by `redrock` and facilitates the evaluation of the quality of the best-fit. This software displays the best-fit template but also the list of fits by order of increasing  $\chi^2$ . The display of these second best-fits allows to detect misclassifications more easily. Finally, the major emission and absorption lines are displayed according to the redshift. An example of QSO spectrum with its `redrock` template is given in Fig. 4.16. By combining the VI results from several people ("VIers") on a reference set of spectra, a truth table was created. This table was compared to the results of `redrock` to assess its efficiency. The details of these results are given for the first DESI data in [39, 40].

At the end of the DESI pipeline, the final product is a set of spectra, and redshift catalogs for each object class.

## Bibliography

- [1] Ž. Ivezić, S. M. Kahn, J. A. Tyson, B. Abel, E. Acosta, R. Allsman et al., *LSST: From Science Drivers to Reference Design and Anticipated Data Products*, *The Astrophysical Journal* **873** (2019) 111.
- [2] D. G. York, *The Sloan Digital Sky Survey: Technical Summary*, *The Astronomical Journal* **120** (2000) 1579.
- [3] S. A. Smee, J. E. Gunn, A. Uomoto, N. Roe, D. Schlegel, C. M. Rockosi et al., *The Multi-object, Fiber-fed Spectrographs for the Sloan Digital Sky Survey and the Baryon Oscillation Spectroscopic Survey*, *The Astronomical Journal* **146** (2013) 32.
- [4] J. E. Gunn, W. A. Siegmund, E. J. Mannery, R. E. Owen, C. L. Hull, R. F. Leger et al., *The 2.5 m Telescope of the Sloan Digital Sky Survey*, *The Astronomical Journal* **131** (2006) 2332.
- [5] D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol et al., *Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies*, *The Astrophysical Journal* **633** (2005) 560.
- [6] D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. A. Prieto, S. F. Anderson et al., *SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way Galaxy, and Extra-Solar Planetary Systems*, *The Astronomical Journal* **142** (2011) 72.
- [7] K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, S. Bailey et al., *The Baryon Oscillation Spectroscopic Survey of SDSS-III*, *The Astronomical Journal* **145** (2013) 10.
- [8] N. G. Busca, T. Delubac, J. Rich, S. Bailey, A. Font-Ribera, D. Kirkby et al., *Baryon Acoustic Oscillations in the Ly- $\alpha$  forest of BOSS quasars*, *Astronomy & Astrophysics* **552** (2013) A96.
- [9] M. R. Blanton, M. A. Bershadsky, B. Abolfathi, F. D. Albareti, C. A. Prieto, A. Almeida et al., *Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe*, *arXiv:1703.00052 [astro-ph]* (2017) .
- [10] R. Ahumada, C. A. Prieto, A. Almeida, F. Anders, S. F. Anderson, B. H. Andrews et al., *The Sixteenth Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra*, *arXiv:1912.02905 [astro-ph]* (2020) .
- [11] K. S. Dawson, J.-P. Kneib, W. J. Percival, S. Alam, F. D. Albareti, S. F. Anderson et al., *The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data*, *The Astronomical Journal* **151** (2016) 44.
- [12] M. Ata, F. Baumgarten, J. Bautista, F. Beutler, D. Bizyaev, M. R. Blanton et al., *The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: first measurement of baryon acoustic oscillations between redshift 0.8 and 2.2*, *Monthly Notices of the Royal Astronomical Society* **473** (2018) 4773.

- [13] A. de Mattia, V. Ruhlmann-Kleider, A. Raichoor, A. J. Ross, A. Tamone, C. Zhao et al., *The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the emission line galaxy sample from the anisotropic power spectrum between redshift 0.6 and 1.1*, *Monthly Notices of the Royal Astronomical Society* **501** (2021) 5616.
- [14] EBOSS collaboration, *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological Implications from two Decades of Spectroscopic Surveys at the Apache Point observatory*, *Physical Review D* **103** (2021) 083533.
- [15] A. S. Bolton, D. J. Schlegel, E. Aubourg, S. Bailey, V. Bhardwaj, J. R. Brownstein et al., *Spectral Classification and Redshift Measurement for the SDSS-III Baryon Oscillation Spectroscopic Survey*, *The Astronomical Journal* **144** (2012) 144.
- [16] I. Pâris, P. Petitjean, E. Aubourg, S. Bailey, N. P. Ross, A. D. Myers et al., *The Sloan Digital Sky Survey quasar catalog: ninth data release*, *Astronomy & Astrophysics* **548** (2012) A66.
- [17] DESI collaboration, *The DESI Experiment Part I: Science, Targeting, and Survey Design*, *arXiv:1611.00036 [astro-ph]* (2016) .
- [18] DESI collaboration, *The DESI Experiment Part II: Instrument Design*, *arXiv:1611.00037 [astro-ph]* (2016) .
- [19] C.-A. Claveau, *Construction d'un spectrographe et recherche de quasars pour le projet d'étude de l'énergie noire, DESI*, phdthesis, Université Paris Saclay (COmUE), Oct., 2019.
- [20] R. Laureijs, J. Amiaux, S. Arduini, J.-L. Auguères, J. Brinchmann, R. Cole et al., *Euclid Definition Study Report*, *arXiv:1110.3193 [astro-ph]* (2011) .
- [21] G. D. Racca, R. Laureijs, L. Stagnaro, J. C. Salvignol, J. L. Alvarez, G. S. Criado et al., *The Euclid mission design*, *arXiv:1610.05508 [astro-ph]* (2016) 990400.
- [22] A. D. Hincks, V. Acquaviva, P. A. R. Ade, P. Aguirre, M. Amiri, J. W. Appel et al., *The atacama cosmology telescope (ACT): Beam profiles and first SZ cluster maps*, *The Astrophysical Journal Supplement Series* **191** (2010) 423.
- [23] E. Chaussidon, C. Yèche, N. Palanque-Delabrouille, D. M. Alexander, J. Yang, S. Ahlen et al., *Target Selection and Validation of DESI Quasars*, Aug., 2022. 10.48550/arXiv.2208.08511.
- [24] DESI collaboration, *Validation of the Scientific Program for the Dark Energy Spectroscopic Instrument*, in preparation (2022).
- [25] A. D. Myers, J. Moustakas, S. Bailey, B. A. Weaver, A. P. Cooper, J. E. Forero-Romero et al., *The Target Selection Pipeline for the Dark Energy Spectroscopic Instrument*, Aug., 2022. 10.48550/arXiv.2208.08518.
- [26] C. Hahn, M. J. Wilson, O. Ruiz-Macias, S. Cole, D. H. Weinberg, J. Moustakas et al., *DESI Bright Galaxy Survey: Final Target Selection, Design, and Validation*, Aug., 2022. 10.48550/arXiv.2208.08512.

- [27] R. Zhou, B. Dey, J. A. Newman, D. J. Eisenstein, K. Dawson, S. Bailey et al., *Target Selection and Validation of DESI Luminous Red Galaxies*, Aug., 2022. 10.48550/arXiv.2208.08515.
- [28] A. Raichoor, J. Moustakas, J. A. Newman, T. Karim, S. Ahlen, S. Alam et al., *Target Selection and Validation of DESI Emission Line Galaxies*, Aug., 2022. 10.48550/arXiv.2208.08513.
- [29] B. Abareshi, J. Aguilar, S. Ahlen, S. Alam, D. M. Alexander, R. Alfarsy et al., *Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument*, Tech. Rep. arXiv:2205.10939, arXiv, May, 2022. 10.48550/arXiv.2205.10939.
- [30] M. Warner, R. Cantarutti, G. Schumacher, E. Mondaca, O. Estay, M. Martinez et al., *A modern approach to upgrading the telescope control system of the CTIO Blanco 4-m telescope*, in *Software and Cyberinfrastructure for Astronomy II*, vol. 8451, pp. 297–305, SPIE, Sept., 2012, DOI.
- [31] D. Sprayberry, P. Dunlop, M. Evatt, L. Reddell, S. Gott, J. R. George et al., *Modernization of the Mayall Telescope control system: design, implementation, and performance*, in *Ground-based and Airborne Telescopes VI*, vol. 9906, pp. 77–88, SPIE, Aug., 2016, DOI.
- [32] T. N. Miller, R. W. Besuner, M. E. Levi, M. Lampton, P. Jelinsky, H. Heetderks et al., *Fabrication of the DESI corrector lenses*, in *Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation III*, vol. 10706, pp. 256–264, SPIE, July, 2018, DOI.
- [33] D. Brooks, P. Doel, R. Besuner, B. Flaugher, G. Gallo, G. Gutierrez et al., *The alignment and assembly of the DESI prime focus corrector*, in *Ground-based and Airborne Instrumentation for Astronomy VI*, vol. 9908, pp. 2483–2490, SPIE, Aug., 2016, DOI.
- [34] J. H. Silber, P. Fagrelus, K. Fanning, M. Schubnell, J. N. Aguilar, S. Ahlen et al., *The Robotic Multi-Object Focal Plane System of the Dark Energy Spectroscopic Instrument (DESI)*, Tech. Rep. arXiv:2205.09014, arXiv, May, 2022. 10.48550/arXiv.2205.09014.
- [35] J. Edelstein, P. Jelinsky, M. Levi, G. Tarle and D. Brooks, *The DESI spectrograph system and production*, in *Ground-based and Airborne Instrumentation for Astronomy VII*, vol. 10702, pp. 2255–2272, SPIE, July, 2018, DOI.
- [36] J. Guy, S. Bailey, A. Kremin, S. Alam, C. A. Prieto, S. BenZvi et al., *The Spectroscopic Data Processing Pipeline for the Dark Energy Spectroscopic Instrument*, Sept., 2022. 10.48550/arXiv.2209.14482.
- [37] A. S. Bolton and D. J. Schlegel, *Spectro-Perfectionism: An Algorithmic Framework for Photon Noise-Limited Extraction of Optical Fiber Spectroscopy*, *Publications of the Astronomical Society of the Pacific* (2010) 100119133735095.
- [38] DESI collaboration, *Redrock spectral classification and redshift fitting*, in preparation (2022).

- [39] D. M. Alexander, T. M. Davis, E. Chaussidon, V. A. Fawcett, A. X. Gonzalez-Morales, T.-W. Lan et al., *The DESI Survey Validation: Results from Visual Inspection of the Quasar Survey Spectra*, Aug., 2022. 10.48550/arXiv.2208.08517.
- [40] T.-W. Lan, R. Tojeiro, E. Armengaud, J. X. Prochaska, T. M. Davis, D. M. Alexander et al., *The DESI Survey Validation: Results from Visual Inspection of Bright Galaxies, Luminous Red Galaxies, and Emission Line Galaxies*, Aug., 2022. 10.48550/arXiv.2208.08516.

# 5

---

## One-dimensional Lyman- $\alpha$ power spectrum with DESI

---

*“The power of a controversial research result is enormous. It attracts physicists like picnics attract ants.”*

– Kip Thorne, *Black Holes & Time Warps: Einstein’s Outrageous Legacy*, 1994

---

**Contents**

<b>5.1</b>	<b>One-dimensional power spectrum pipeline</b>	<b>164</b>
5.1.1	Catalog creation	164
5.1.2	Lyman- $\alpha$ contrast extraction	167
5.1.3	Power spectrum estimation	168
5.1.4	Parameters used for one-dimensional power spectrum calculation	170
<b>5.2</b>	<b>Systematics measurement on DESI data</b>	<b>172</b>
5.2.1	Noise estimation	172
5.2.2	Side-band power spectrum	177
5.2.3	Sky emission lines	180
<b>5.3</b>	<b>Characterization on mocks</b>	<b>183</b>
5.3.1	Pixel masking	183
5.3.2	Continuum fitting	185
<b>5.4</b>	<b>One-dimensional Lyman-<math>\alpha</math> power spectrum measurement on DESI data</b>	<b>187</b>
5.4.1	Determination of uncertainties	188
5.4.2	Measurement	191
5.4.3	Summary and prospects	192
	<b>Bibliography</b>	<b>194</b>

---

THE small-scale distribution of matter is imprinted in the fluctuations of the Ly $\alpha$  forest along the line-of-sight. Measurements of the one-dimensional Ly $\alpha$  power spectrum ( $P_{1D,\alpha}$ ) are a powerful tool to probe the cosmological clustering of matter together with the thermal state of the IGM. They can be performed from optical quasar spectra between  $z = 2$  and  $z \sim 5$ . As the small-scale matter distribution is sensitive to the absolute mass scale of neutrinos and WDM models (see Sec. 1.3.4),  $P_{1D,\alpha}$  is an optimal observable to set constraints on these parameters [1].

Measuring  $P_{1D,\alpha}$  requires first to extract the fluctuations of the Ly $\alpha$  forest after renormalizing by the quasar continuum. From there, several methods can be used to compute  $P_{1D,\alpha}$ . The most straightforward is to compute the fast Fourier transform (FFT) of the Ly $\alpha$  contrasts. This was used in BOSS and eBOSS [2, 3] analyses. The one-dimensional power spectrum can also be measured using a quadratic maximum likelihood estimator (QMLE). This more complex method has already been used for moderate resolution observations [2, 4] and more recently on high-resolution data [5]. In this thesis, we use an FFT estimation, as presented in Sec. 5.1.

Our purpose is to compute  $P_{1D,\alpha}$  from the first DESI data, following the same methodology as in the latest eBOSS measurement [3]. This will facilitate the comparison between eBOSS and DESI.  $P_{1D,\alpha}$  is an observable sensitive to instrumental properties such as noise and spectral resolution: as the telescopes used and the data are very different, it is essential to characterize the DESI instrument as shown in Sec. 5.2. We improved the algorithms and methods used in [3] and adapted them to the DESI data.

To test the different steps of the  $P_{1D,\alpha}$  pipeline and in order to apply adequate corrections to our measurement, we used the specially designed mocks presented in Sec. 3.1.2. In particular, these mocks allowed us to evaluate in Sec. 5.3 the impact of pixel masking and biases introduced by the  $P_{1D,\alpha}$  pipeline.

Finally, the measurement using the first DESI data is presented in Sec. 5.4, along with its associated statistical and systematic uncertainties. With only a minor fraction of the Ly $\alpha$  forest available at the end of the DESI survey, this measurement already improves over eBOSS on small scales (high wavenumber).

For this analysis, I developed the python package `pidesi`<sup>1</sup>. All algorithms used in this chapter are documented in this public package.

---

1. <https://github.com/corentinravoux/pidesi/>

## 5.1 One-dimensional power spectrum pipeline

The objective of this chapter is to measure  $P_{1D,\alpha}$  from spectra computed by the DESI pipeline presented in Sec. 4.3.4. The first step consists in creating the QSO, HCD, sky line, and BAL catalogs (Sec. 5.1.1). Along with DESI spectra, all these catalogs are used to compute the Ly $\alpha$  contrast (Sec. 5.1.2). Finally,  $P_{1D,\alpha}$  is computed using an FFT method for all the selected Ly $\alpha$  forest samples (Sec. 5.1.3). A schematic representation of the pipeline used in this analysis is given in Fig. 5.1.

### 5.1.1 Catalog creation

Several catalogs are necessary to start processing the Ly $\alpha$  forest. This pre-processing part is shared with other Ly $\alpha$  and QSO-related measurements.

#### Quasars

The DESI pipeline categorizes the observed spectra with the `redrock` software and gives a redshift estimate as explained in Sec. 4.3.4.3. In order to maximize the completeness and purity of the QSO catalog used, a broad MgII line finder `mgii_afterburner`<sup>2</sup> and a machine learning classifier `QuasarNET`<sup>3</sup> [7, 8] are runned after `redrock`. Both post-processing programs are run on all objects targeted by DESI.

During the visual inspection of SV DESI data, a low completeness problem of `redrock` was found for  $z \lesssim 1.6$ . An example of spectrum not identified as a quasar is given in Fig. 5.2. `Redrock` returns a galaxy classification for this spectrum because of the presence of narrow peaks at large wavelengths. It was decided within the DESI collaboration to classify this spectrum as a quasar. A modification of `redrock` to take into account this type of QSO would impact all other targets. Therefore, I implemented and tested `mgii_afterburner` whose role was to test the presence of a MgII broad line on the spectra categorized as galaxies by `redrock`.

The `mgii_afterburner` only works for objects whose redshift is correctly determined by `redrock`. It fits a Gaussian function on a wavelength range centered around the MgII peak  $\lambda_{\text{MgII}} = \lambda_{\text{MgII,rest}}(1 + z)$ :

$$G(A, B, \sigma, \lambda) = A \exp\left(\frac{-(\lambda - \lambda_{\text{MgII}})^2}{2\sigma^2}\right) + B. \quad (5.1)$$

For each spectrum, the  $\chi^2$  of the `mgii_afterburner` is estimated over the same wavelength range such that:

$$\chi_{\text{mgii}}^2(A, B, \sigma) = \sum_{\lambda \in [\lambda_{\text{MgII}} \pm \delta\lambda]} \left(\frac{f(\lambda) - G(A, B, \sigma, \lambda)}{\sigma_f(\lambda)}\right)^2, \quad (5.2)$$

where  $\delta\lambda = 125 \text{ \AA}$  defines the wavelength range of the fitting procedure. Similarly, we calculate  $\chi_{\text{redrock}}^2$  by replacing the  $G$  function with the best-fit model of `redrock`. To select spectra with a MgII peak considered as broad, I use the parameters  $\Delta\chi^2 = \chi_{\text{mgii}}^2 - \chi_{\text{redrock}}^2$ ,  $\sigma$ ,  $A$ , and  $\sigma_A/A$  where  $\sigma_A$  is the fitting error on  $A$ . I optimized these constraints to improve the completeness while keeping the purity of a quasar sample processed by `redrock` and compared to visual inspection.

2. [https://github.com/desihub/desispec/blob/master/py/desispec/mgii\\_afterburner.py](https://github.com/desihub/desispec/blob/master/py/desispec/mgii_afterburner.py)

3. Numpy implementation of QuasarNET: <https://github.com/desihub/QuasarNP>

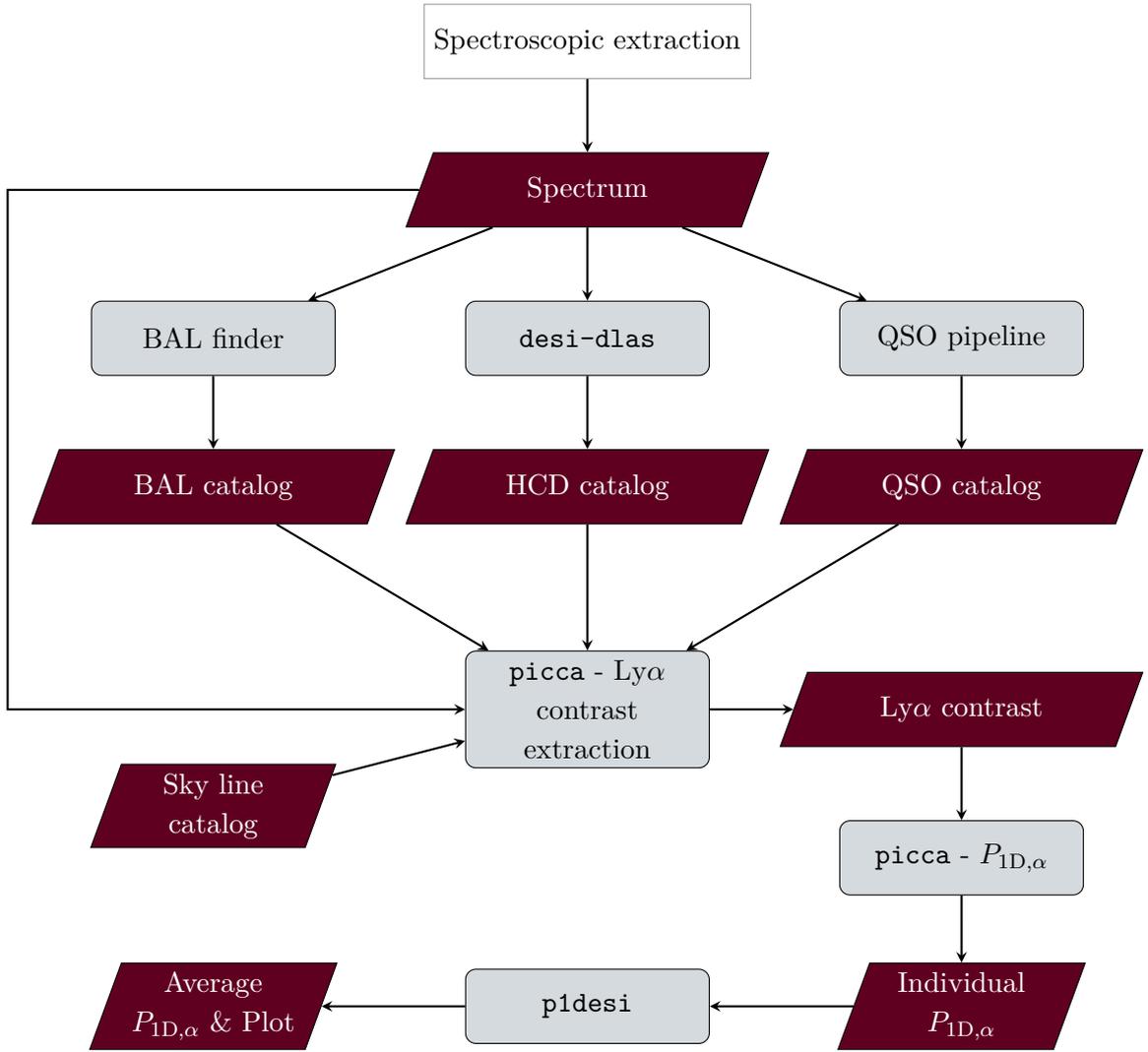


FIGURE 5.1 – Pipeline used for the calculation of Ly $\alpha$  contrasts and one-dimensional Ly $\alpha$  power spectrum. The details concerning the QSO pipeline are given in [6].

This study was carried out on the SV data set, in [6, 9]. The parameters I determined are the following:

$$\begin{aligned}
 10 < \sigma < 200 \text{ \AA}, \\
 \Delta\chi^2 > 16, \\
 \sigma_A/A > 3, \\
 A > 0.
 \end{aligned}
 \tag{5.3}$$

The object represented in Fig. 5.2 is an example of quasar missed both by `redrock` and visual inspection, which is recovered with `mgii_afterburner`. Some faint QSO spectra missed by `redrock` and `mgii_afterburner` are recovered using `QuasarNET`. This algorithm is a convolutional neural network classifier consisting of four convolution layers, a fully-connected layer, and individual line finder units. Each unit comprises a fully-connected layer trained to identify a particular emission line. In the case of the QSO catalog creation, the spectra are the input of `QuasarNET`, and six lines (Ly $\alpha$ , C $\text{IV}$ , C $\text{II}$ , Mg $\text{II}$ , H $\alpha$ , H $\beta$ ) are used for classification. An object is classified as a QSO if at least one of the six confidence probabilities is larger than 50 %.

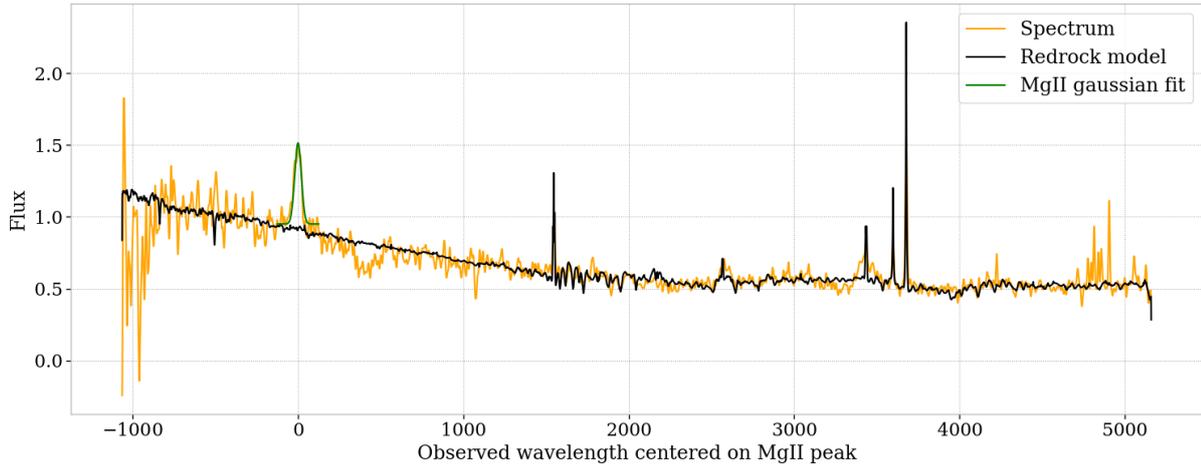


FIGURE 5.2 – Example of a DESI quasar spectrum in yellow ( $z = 0.67$  and TARGETID 39633351631439695) showing a broad MgII emission together with narrow, galactic emission lines. The fit made by the `mgii_afterburner` is shown in green and the `redrock` best-fit model in black. The pipeline reduction used is Blanc.

`QuasarNET` is efficient at categorizing missed quasars but does not provide an accurate redshift in comparison to visual inspection. The redshift of retrieved quasars is obtained by running `redrock` again with a `QuasarNET` redshift prior, considering only QSO templates.

`Mgii_afterburner` is run on all objects identified differently than QSO by `redrock`, without prior from `QuasarNET`. The `QuasarNET` afterburner is run both for objects identified as QSO by `redrock` and objects simultaneously rejected by `redrock` and `mgii_afterburner`. The details on this QSO catalog pipeline and the tests carried out is published in [6]. The addition of afterburners increased the completeness of the QSO catalog from 80 % to 98 % without notably degrading its purity. Here, the completeness is defined as the number of QSO detected by the pipeline divided by the number visually inspected as QSO. The purity measures the number of objects detected as QSO by the pipeline which are not visually inspected as a QSO.

### Additional catalogs

BAL QSOs (see Sec. 2.1.3) are searched for in quasar spectra with a  $\chi^2$  minimizer algorithm that looks for blueshifted CIV or SiIV absorptions in an unabsorbed quasar model. The fit is performed for rest-frame wavelengths between 1260 and 2400 Å. The algorithm measures the balnicity index defined by Eqn. 2.21. BAL quasar spectra are not used in all the  $P_{1D,\alpha}$  pipeline.

High Column Density (HCD) objects, such as DLAs, also contaminate the Ly $\alpha$  contrast. We use the HCD finder `desi-dlas`<sup>4</sup> [10] to remove this contamination. It is a convolutional neural network (CNN) algorithm trained with SDSS spectra to identify candidate HCD for rest-frame wavelengths between 900 and 1346 Å. It returns HCD object locations in the spectra as well as their HI column density (defined by Eqn. 2.7) and a confidence parameter. Although HCD objects by themselves constitute tracers of the matter distribution, they have an extended impact on the observed spectra, as detailed in Sec. 2.1.3, and can create additional correlations on several pixels in the  $P_{1D,\alpha}$  measurement. Therefore, we choose to mask the affected regions of the spectra. We remove the pixels where the HCD-induced absorption is larger than 20 %.

4. <https://github.com/cosmodesi/desi-dlas>

addition, the absorption in the Lorentzian-profile damping wings that remain after the cut are corrected with a Voigt profile following [11, 12].

Finally, the line emissions from the atmosphere called sky lines increase the pixel noise. Following [2, 3], we mask some sky lines during the Ly $\alpha$  contrast extraction and the  $P_{1D,\alpha}$  calculation. The creation of the sky lines catalog is presented in the DESI framework in Sec. 5.2.3.

### 5.1.2 Lyman- $\alpha$ contrast extraction

The Ly $\alpha$  contrast  $\delta_F(\lambda)$  is given in our framework by Eqn. 2.22. The **Ly $\alpha$  contrast extraction**, otherwise called **continuum fitting**, consists in determining from the quasar flux  $f(\lambda)$ , the continuum quasar flux  $C_q(\lambda, z_q)$  together with the average fraction of transmitted flux  $\bar{F}(\lambda)$ . Following [3, 13], this procedure is performed with `picca`<sup>5</sup> [14]. This software was built for BOSS/eBOSS surveys and adapted to DESI data processing.

For moderate-resolution observations like DESI, low-SNR spectra make it difficult to measure the "upper envelope" of the measured flux, i.e.  $C_q(\lambda, z_q)$ , accurately. Thus, the continuum fitting is performed in `picca` by actually measuring the product  $C_q(\lambda, z_q)\bar{F}(\lambda)$ . Each quasar continuum is modeled as the product of a universal continuum  $C(\lambda_{\text{rf}})$  common to all quasars, and a first-order polynomial term in wavelength:

$$C_q(\lambda, z_q) = (a_q + b_q\Lambda) C\left(\lambda_{\text{rf}} = \frac{\lambda}{(1 + z_q)}\right), \quad (5.4)$$

where  $a_q$  and  $b_q$  are quasar-dependent normalization terms. The variable  $\Lambda$  is defined depending on the wavelength binning:

$$\Lambda = \begin{cases} \log(\lambda), & \text{logarithmic binning (SDSS)}, \\ \lambda, & \text{linear binning (DESI)}. \end{cases} \quad (5.5)$$

The  $a_q$  and  $b_q$  parameters along with the  $C$  function are determined for each quasar by minimizing the following likelihood:

$$\mathcal{L} = - \sum_i \frac{[f_i - \bar{F}(\lambda_i)C_q(\lambda_i, z_q, a_q, b_q)]^2}{\sigma_q^2(\lambda_i)} - \ln [\sigma_q^2(\lambda_i)], \quad (5.6)$$

where the sum is run over all the forest pixels of the quasar  $q$ . The standard deviation estimator of the flux  $f$ , noted  $\sigma_q$  is the sum of several contributions:

$$\frac{\sigma_q^2(\lambda)}{(\bar{F}(\lambda)C_q(\lambda))^2} = \eta(\lambda) \frac{\sigma_{\text{pip},q}^2(\lambda)}{(\bar{F}(\lambda)C_q(\lambda))^2} + \sigma_{\text{iss}}^2(\lambda) + \epsilon(\lambda) \frac{(\bar{F}(\lambda)C_q(\lambda))^2}{\sigma_{\text{pip},q}^2(\lambda)}. \quad (5.7)$$

Here,  $\sigma_{\text{pip},q}$  is the noise given by the spectral extraction pipeline (detailed in Sec. 4.2.3 for SDSS and Sec. 4.3.4.2 for DESI). A wavelength-dependent term  $\eta$  is added to correct for potential noise underestimations. The large-scale structure variance,  $\sigma_{\text{iss}}^2$ , is the intrinsic variance of the flux-transmission field and gives minimal floor variance. Finally, the third factor controlled by  $\epsilon$  is included to compensate the observed increase of the variance at high SNR, due to the quasar diversity or variability.

---

5. <https://github.com/igmhub/picca>

The functions  $C$ ,  $\eta$ ,  $\sigma_{\text{ISS}}$ , and  $\epsilon$  are computed iteratively. In particular,  $C$  assumes no parameterization and is estimated from the average of all spectra. The  $\eta$ ,  $\sigma_{\text{ISS}}$ , and  $\epsilon$  functions are computed by fitting the variance of the Ly $\alpha$  contrast at each step.

During the entire fitting procedure, pixels that are hidden due to an HCD object or a sky line are not considered in the fit.

To characterize the noise level of a Ly $\alpha$  forest, I define the **average signal-to-noise ratio** of a forest

$$\overline{\text{SNR}} = \left\langle \frac{f(\lambda)}{\sigma_{\text{pip,q}}(\lambda)} \right\rangle_{\lambda}. \quad (5.8)$$

The standard deviation  $\sigma_{\delta_F}$  associated with the Ly $\alpha$  contrast is defined by

$$\sigma_{\delta_F}(\lambda) = \eta(\lambda) \frac{\sigma_{\text{pip,q}}(\lambda)}{F(\lambda)C_q(\lambda)}. \quad (5.9)$$

The output of the continuum fitting is the set  $(\delta_F, \sigma_{\delta_F})$  for all selected quasars.

### 5.1.3 Power spectrum estimation

#### Ly $\alpha$ contrast decomposition

Several effects contribute to the measured Ly $\alpha$  contrast. I separate these different contributions such that:

$$\delta_F(\lambda) = \delta_{\text{Ly}\alpha}(\lambda) + \delta_{\text{others}}(\lambda) + \delta_{\text{noise}}(\lambda). \quad (5.10)$$

In this decomposition,  $\delta_{\text{Ly}\alpha}$  corresponds to the fluctuations caused by the Ly $\alpha$  absorption, which is the signal we want to measure. The absorptions by other metals drive the  $\delta_{\text{others}}$  fluctuations, and  $\delta_{\text{noise}}$  corresponds to noise fluctuations.

In the metals contributing to  $\delta_{\text{others}}(\lambda)$ , we separate the contribution of lines close in wavelength to the Ly $\alpha$  emission and the rest of metals: the SiII and SiIII elements possess rest-frame line absorptions at  $\lambda_{\text{SiII}} = 1190$  and  $1193$  Å, and  $\lambda_{\text{SiIII}} = 1206.50$  Å. These wavelengths are close enough from  $\lambda_{\text{Ly}\alpha} = 1215.67$  Å so that the correlated Ly $\alpha$  and SiII/SiIII absorptions from the same absorbing clouds is observable with  $P_{\text{1D},\alpha}$ : they create oscillatory patterns on  $P_{\text{1D},\alpha}$ , characterized by a contrast that we will note  $\delta_{\text{SiII/SiIII}}$ . By adopting a method suggested in [4] and used in all other analyses, the absorption contributions of the SiII and SiIII are kept in the  $P_{\text{1D},\alpha}$  observable, as they can be taken into account analytically during its modeling. I group the absorption of all the other metals in a contrast noted  $\delta_{\text{metals}}$ .

The measured  $\delta_F$  is a pixelized signal processed by the spectroscopic pipeline. We take into account the pixelization of  $\delta_F$  by convolving a top-hat function  $\Pi$  and the finite spectrograph resolution using a function  $W(\lambda, \mathbf{R})$ . The measured Ly $\alpha$  contrast thus reads:

$$\delta_F(\lambda) = \left( \delta_{\text{Ly}\alpha}(\lambda) + \delta_{\text{SiII/SiIII}}(\lambda) + \delta_{\text{metals}}(\lambda) \right) \otimes W(\lambda, \mathbf{R}) \otimes \Pi(\lambda, \Delta\lambda_{\text{pix}}) + \delta_{\text{noise}}(\lambda), \quad (5.11)$$

where  $\Delta\lambda_{\text{pix}}$  is the separation between each pixel. Here, we assumed that the impact of the noise and the resolution matrix are decorrelated, i.e., that the noise contrast is not affected by the instrumental effects of pixelization and resolution.

Based on this decomposition, we want to measure the one-dimension Ly $\alpha$  power spectrum, theoretically defined by Eqn. 2.27. I estimate  $P_{1D,\alpha}$  by applying a **fast Fourier transform (FFT)** algorithm on the Ly $\alpha$  contrasts. This method is the most straightforward and has been applied in previous measurements [2, 3, 4, 15]. The FFT estimator is implemented in `picca`<sup>5</sup> [14] for both SDSS and DESI data. For each Ly $\alpha$  forest, we define its **raw power spectrum**<sup>6</sup> based on the Fourier transform of its Ly $\alpha$  contrast:

$$P_{\text{raw}}(k) = |\delta_F(k)|^2. \quad (5.12)$$

Applying a Fourier transform on Eqn. 5.11, the raw power spectrum is expressed by

$$P_{\text{raw}}(k) = \left( P_{\text{Ly}\alpha}(k) + P_{\text{Ly}\alpha\text{-SiII/SiIII}}(k) + P_{\text{metals}}(k) \right) \cdot W^2(k, \mathbf{R}) \cdot \text{sinc}^2\left(\frac{k\Delta\lambda_{\text{pix}}}{2}\right) + P_{\text{noise}}(k). \quad (5.13)$$

In this decomposition,  $P_{\text{Ly}\alpha}$ ,  $P_{\text{metals}}$ , and  $P_{\text{noise}}$  are the power spectra associated to the contrasts  $\delta_{\text{Ly}\alpha}$ ,  $\delta_{\text{metals}}$ , and  $\delta_{\text{noise}}$  with the same definition as Eqn. 5.12. We assumed here that the SiII and SiIII power spectra are negligible and that the cross-correlation terms between these contrasts are null. The only non-neglected cross-term is  $P_{\text{Ly}\alpha\text{-SiII/SiIII}}(k) = 2 \left| \delta_{\text{Ly}\alpha}(k) \delta_{\text{SiII/SiIII}}(k) \right|$ , which corresponds to the correlated absorptions of Ly $\alpha$  with either SiII or SiIII. The oscillations induced by this term have a wavenumber  $2\pi/(\lambda_{\text{Ly}\alpha} - \lambda_{\text{SiII/SiIII}})$  when  $k$  is expressed in  $\text{\AA}^{-1}$ .

### Pixelization and resolution in DESI

For SDSS  $P_{1D,\alpha}$  measurements [2, 3, 4, 15], the pixelization is log binned in observed wavelength, making it suitable to express the power spectrum in Hubble velocity unit  $v$ , because  $\Delta v \propto \Delta\lambda/\lambda = \Delta \log(\lambda)$ . The constant SDSS pixel width is  $\Delta v_{\text{pix}} = 69 \text{ s}\cdot\text{km}^{-1}$ . The resolution correction is modeled by a Gaussian function  $W(k, \Delta\lambda) = \exp(-0.5(k\Delta\lambda)^2)$ , where  $\Delta\lambda$  is the spectrograph resolution power.

With DESI, as indicated in Sec. 4.3.4, spectra are linearly binned with a constant pixel width  $\Delta\lambda_{\text{pix}} = 0.8 \text{ \AA}$ , so that wavenumbers are expressed in  $\text{\AA}^{-1}$ . The maximal measurable wavenumber follows the Nyquist-Shannon limit:  $k_{\text{max}} = k_{\text{Nyq}} = \pi/\Delta\lambda_{\text{pix}} \simeq 3.92 \text{ \AA}^{-1}$ . In practice, this limit is also impacted by the spectrograph resolution as shown in 5.2.1.

The DESI spectrograph resolution is entirely defined by the resolution matrix (Eqn. 4.5), which is used to link the extracted flux  $\vec{f}_c$  to the flux measured on the CCD  $f$  such that  $\vec{f}_c = \mathbf{R}\vec{f}$ . Considering only the impact of resolution and pixelization, we need to find the function  $W$  which verifies  $\delta_F(k) = \delta_{\text{Ly}\alpha}(k) \cdot W(k, \mathbf{R}) \cdot \text{sinc}(k\Delta\lambda_{\text{pix}}/2)$ . One must take care of the fact that the resolution matrix  $\mathbf{R}$  is itself pixelized, and even doubly-pixelized over both columns and rows. Consequently, it can be expressed as

$$\mathbf{R}(\lambda) = W(\lambda, \mathbf{R}) \otimes \Lambda(\lambda, \Delta\lambda_{\text{pix}}), \quad (5.14)$$

where  $\Lambda = \Pi \otimes \Pi$  is a triangle function, and  $W$  is the binning-independent spectroscopic resolution kernel. By applying a Fourier transform, we can express the resolution correction in Fourier space:

$$W(k, \mathbf{R}) = \frac{\mathbf{R}(k)}{\text{sinc}^2\left(\frac{k\Delta\lambda_{\text{pix}}}{2}\right)}. \quad (5.15)$$

---

6. This is not a proper power spectrum, strictly speaking, as it does not correspond to an ensemble average.

Finally, considering only resolution and pixelization effects, the Ly $\alpha$  power spectrum is given by

$$P_{\text{Ly}\alpha}(k) = |\delta_{\text{Ly}\alpha}(k)|^2 = P_{\text{raw}}(k) \cdot \mathbf{R}^{-2}(k) \cdot \text{sinc}^2\left(\frac{k\Delta\lambda_{\text{pix}}}{2}\right). \quad (5.16)$$

### FFT estimator

The FFT estimator for the one-dimensional power spectrum is computed as an ensemble average over all available Ly $\alpha$  forest in the measurement sample. It is designed to match the sum of Ly $\alpha$  and Ly $\alpha$ -SiII/SiIII power spectra in Eqn. 5.13, so that we define

$$P_{\text{1D},\alpha}(k) = \left\langle P_{\text{Ly}\alpha}(k) + P_{\text{Ly}\alpha\text{-SiII/SiIII}}(k) \right\rangle, \quad (5.17)$$

where  $\langle \cdot \rangle$  denotes the ensemble average over all the Ly $\alpha$  forests used for  $P_{\text{1D},\alpha}$  calculation.

From Eqn. 5.16, adding the metals and noise power spectra, the estimator of  $P_{\text{1D},\alpha}$  is defined by

$$P_{\text{1D},\alpha}(k) = \left\langle [P_{\text{raw}}(k) - P_{\text{noise}}(k)] \cdot \mathbf{R}^{-2}(k) \cdot \text{sinc}^2\left(\frac{k\Delta\lambda_{\text{pix}}}{2}\right) \right\rangle - P_{\text{metals}}(k). \quad (5.18)$$

The calculation of the average ensemble is done with the `p1desi`<sup>1</sup> python package.

## 5.1.4 Parameters used for one-dimensional power spectrum calculation

### Ly $\alpha$ contrast extraction

Only Ly $\alpha$  forests with an average signal-to-noise ratio  $\overline{\text{SNR}}$  (see Eqn. 5.8) larger than 1 are used in the continuum fitting procedure.

The Ly $\alpha$  contrast extraction is restricted to pixels for which the observed wavelength is between 3500 and 7600 Å. We also select the rest-frame wavelength in the range  $1040 < \lambda_{\text{rf}} < 1200$  Å, so that the measured contrasts are dominated by the Ly $\alpha$  forest. In particular, we do not attempt to exploit the Ly $\beta$  forest ( $\lambda_{\text{Ly}\beta} < 1025.72$  Å in the rest-frame). The cut  $\lambda_{\text{RF}} < 1200$  Å implies that for a given line-of-sight, the closest selected pixel is located at a minimum redshift separation of  $\Delta z = 0.041$  from the associated quasar. This decision mitigates most of the proximity effect: close to a quasar, the proportion of neutral hydrogen is indeed influenced by the large UV radiation background [16]. It also facilitates the continuum fitting.

In contrast to BAO analysis, we want all spectra pixels to contribute equally to the continuum fitting and the FFT computation for  $P_{\text{1D},\alpha}$  measurement. Therefore, I impose SNR-independent weights in the continuum fitting procedure, which means that in the definition of the flux standard deviation (Eqn. 5.7),  $\eta$  and  $\epsilon$  parameters are set to zero, and  $\sigma_{\text{1ss}} = 1$  for all wavelengths. In that particular case, the  $\eta$  term is removed from the definition of Ly $\alpha$  contrast standard deviation in Eqn. 5.9.

Additionally, we choose to restrict the polynomial in Eqn. 5.4 to order 0 ( $b_q = 0$ ) to obtain a continuum which depends only on the rest-frame wavelength. The order of the polynomial, as well as the constant fitting weights, are chosen following previous studies [2, 3]. The number of continuum fitting iterations is set to 7 after verification of the continuum convergence.

The spectra are linearly binned in observed wavelength with  $\Delta\lambda_{\text{pix}} = 0.8 \text{ \AA}$ . We choose here to perform a rebinning in the quasar rest frame. This is done to reduce the small variations of the common continuum  $C$ , as we are here using a relatively low-statistics dataset. The rest-frame wavelength array is chosen ten times coarser such that:

$$\Delta\lambda_{\text{pix,rf}} = 10 \frac{\Delta\lambda_{\text{pix}}}{1 + z_{\text{min}}}. \quad (5.19)$$

By taking  $z_{\text{min}} = 2.0$  as the lowest redshift used, we obtain  $\Delta\lambda_{\text{pix,rf}} = 2.67 \text{ \AA}$ . Such a rebinning could be relaxed in future DESI measurements.

### $P_{1D,\alpha}$ FFT calculation

The FFT estimator is impacted by the noise power spectrum, especially at small scales. A second  $\overline{\text{SNR}}$  cut is used to select spectra with a low noise level. This cut is variable for the different analyses performed in this chapter.

For observed wavelength  $\lambda \lesssim 3700 \text{ \AA}$ , the data are very noisy because of the atmospheric absorptions. To minimize the impact of this noise, we remove the pixels for which the observed wavelength is lower than  $3750 \text{ \AA}$ , which corresponds to Ly $\alpha$  absorbers located at  $z = 2.085$ . In the future, with a dedicated study to control the noise at small  $\lambda$ , one could extend the  $P_{1D,\alpha}$  analysis around  $z \sim 2$ .

The  $P_{1D,\alpha}$  measurement is splitted in different redshift bins to take into account its evolution. For a given redshift binning, it is necessary to split the Ly $\alpha$  forest into **sub-forest (or chunks)** which correspond to consecutive and non-overlapping sub-regions of equal length. This procedure also reduces the correlations between the different redshift bins. We chose to cut Ly $\alpha$  forests into three chunks whose rest-frame wavelength boundaries are  $\lambda_{\text{rf}} = 1093.3$  and  $1146.6 \text{ \AA}$ , so that the length of each chunk is  $L_{\text{sub}} = 53.3 \text{ \AA}$ . The sub-forest splitting constraints the minimal accessible wavenumber to  $k_{\text{min}} = 2\pi/L_{\text{sub}} = 0.118 \text{ \AA}^{-1}$ . Each chunk spans at most  $\Delta z = 0.2$  and we choose the same  $\Delta z$  to define the redshift binning for  $P_{1D,\alpha}$ .

Finally, we remove chunks shorter than 75 pixels due to a cut in the UV region or to the presence of a large DLA. We also do not consider the Ly $\alpha$  forests with more than 140 masked pixels.

## 5.2 Systematics measurement on DESI data

The data used for the  $P_{1D,\alpha}$  measurement presented here are SV1 and SV3 quasars, processed with pipeline reduction Fuji, and DA0.2 quasars from the Guadalupe reduction. A description of these observations is provided in Sec. 4.3.1. Only the DESI tiles observed during the dark program are used. The spectra coaddition is realized so that observations of a same quasar located on different overlapping DESI tiles are coadded. In this section, we characterize the impact of noise estimation (Sec. 5.2.1), metal power spectrum (Sec. 5.2.2), and sky emission lines (Sec. 5.2.3) on the  $P_{1D,\alpha}$  measurement. For the noise study, we consider separately the SV1, SV3, and DA0.2 data sets.

### 5.2.1 Noise estimation

We estimate the noise power spectrum using the pipeline noise detailed in Sec. 4.3.4.2 or with an exposure difference method.

#### Pipeline noise

We note  $P_{\text{pipeline}}$  the noise power spectrum derived from the pipeline noise estimator.  $P_{\text{pipeline}}$  is computed from the standard deviation  $\sigma_{\delta_F}$  given by Eqn. 5.9. For each unmasked pixel, a contrast  $\delta_{\text{pipeline}}$  is generated following a normal probability distribution such that:

$$\delta_{\text{pipeline}}(\lambda) \hookrightarrow \mathcal{N}(0, \sigma_{\delta_F}(\lambda)) . \quad (5.20)$$

This procedure is repeated  $N_G$  times ( $N_G = 2500$ ) to obtain a converged noise power spectrum. For each quasar, the associated noise power spectrum is the average of the  $N_G$  noise contrasts after Fourier transformation:

$$P_{\text{pipeline}}(k) = \left\langle |\delta_{\text{pipeline}}|^2 \right\rangle_{N_G} . \quad (5.21)$$

#### Exposure difference noise

Another noise estimation can be done using the difference between exposures of the same quasar, when several exposures are available for the same object. Similarly to the estimation done on the spectra in Sec. 4.3.4, the difference between exposures removes the physical signal, leaving only the fluctuations due to noise. I implemented a noise power spectrum estimator using this principle. I define the difference coadd of a quasar of index  $j$  by separating half of its exposures in the even category ( $N_{\text{even}}$  exposures) and the other half in the odd category ( $N_{\text{odd}}$  exposures) such that:

$$\Delta f_j = \frac{1}{2} \left( \frac{\sum_{k=1}^{N_{\text{even}}} (\mathbb{V}_{\text{pip},k})^{-1} f_k}{\sum_{k=1}^{N_{\text{even}}} (\mathbb{V}_{\text{pip},k})^{-1}} - \frac{\sum_{k=1}^{N_{\text{odd}}} (\mathbb{V}_{\text{pip},k})^{-1} f_k}{\sum_{k=1}^{N_{\text{odd}}} (\mathbb{V}_{\text{pip},k})^{-1}} \right) , \quad (5.22)$$

where  $\mathbb{V}_{\text{pip},k} = \sigma_{\text{pip},k}^2$  is the pipeline variance of the exposure  $k$  for quasar  $j$ . In the case where the total number of exposures is even,  $N_{\text{even}} = N_{\text{odd}}$ . The standard deviation of  $\Delta f_j$  can be calculated from the variances of individual exposures:

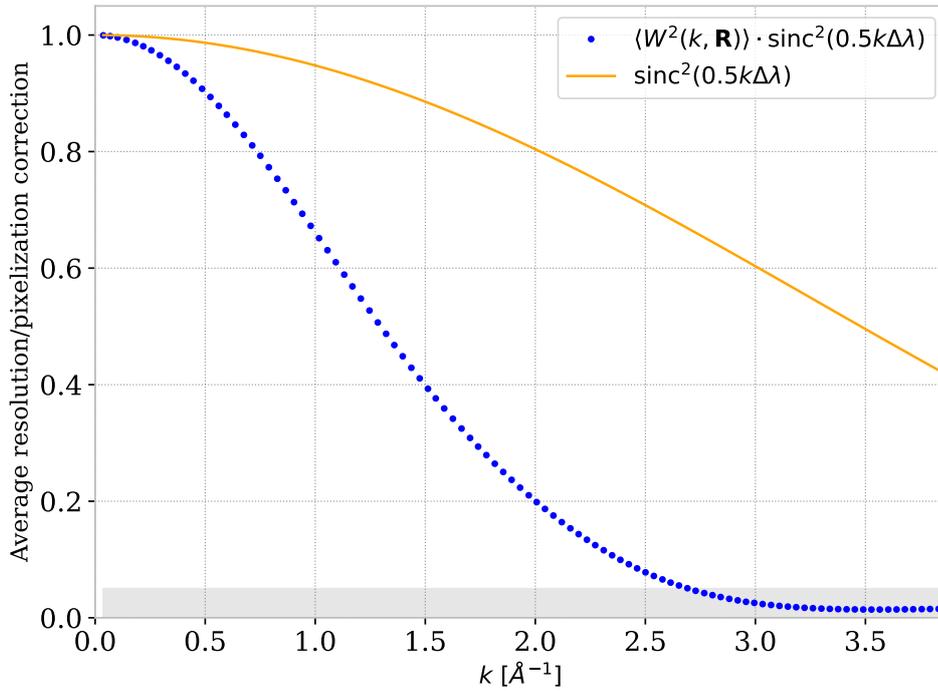


FIGURE 5.3 – Average resolution correction  $\langle W^2(k) \cdot \text{sinc}^2(0.5k\Delta\lambda_{\text{pix}}) \rangle = \langle \mathbf{R}^2(k) \cdot \text{sinc}^{-2}(0.5k\Delta\lambda_{\text{pix}}) \rangle$  using Ly $\alpha$  forest data of SV3 and DA0.2 (blue points), with a  $\overline{\text{SNR}}$  threshold of 3. This resolution correction is not strongly dependent on the redshift range. The mean value over all redshift bins is shown here. As an illustration, the impact of pixelization only, i.e.  $\text{sinc}^2(0.5k\Delta\lambda_{\text{pix}})$ , is represented by a continuous yellow line.

$$\sigma_{\Delta f_j} = \frac{1}{2} \sqrt{\frac{1}{\sum_{k=1}^{N_{\text{even}}} (\mathbb{V}_{\text{pix},k})^{-1}} + \frac{1}{\sum_{k=1}^{N_{\text{odd}}} (\mathbb{V}_{\text{pix},k})^{-1}}}. \quad (5.23)$$

Compared to Sec. 4.3.4.2, this difference coadd was improved: it is unbiased, i.e., of zero average, whatever the values of the sum of the inverse variance for both exposure populations. Finally, this estimator does not necessarily need an even total number of exposures.

To derive an estimator of  $P_{\text{noise}}$ , the variance of  $\Delta f_j$  must be equal to that of the coadded flux (defined by Eqn. 4.7) used to compute  $P_{\text{raw}}$ . The coadded flux variance is given by Eqn. 4.8. To obtain the same variance, I multiply  $\Delta f_j$  by the ratio  $\sigma_{f_j}/\sigma_{\Delta f_j}$ :

$$\Delta f_j^{\text{corr}} = 2 \frac{\frac{1}{\sqrt{\sum_{k=1}^{N_{\text{tot}}} (\mathbb{V}_{\text{pix},k})^{-1}}}}{\sqrt{\frac{1}{\sum_{k=1}^{N_{\text{even}}} (\mathbb{V}_{\text{pix},k})^{-1}} + \frac{1}{\sum_{k=1}^{N_{\text{odd}}} (\mathbb{V}_{\text{pix},k})^{-1}}}} \Delta f_j, \quad (5.24)$$

where  $N_{\text{tot}}$  is the total number of exposure for the quasar  $j$  ( $N_{\text{tot}} = N_{\text{even}} + N_{\text{odd}}$ ).

In SDSS analysis [2, 4, 12], the variance of all exposures for a given object was considered equal. In this case,  $\sigma_{\Delta f_j}$  can be simplified, and a correction was applied only in the case of an odd number of exposures. In the case of a constant exposure variance, the difference coadd in Eqn. 5.24 is equal to the one derived in [2, 3]. Our new estimator corrects the variance for any exposure time. It is essential in the case of DESI SV data, for which the exposure times can be very variable compared to SDSS.

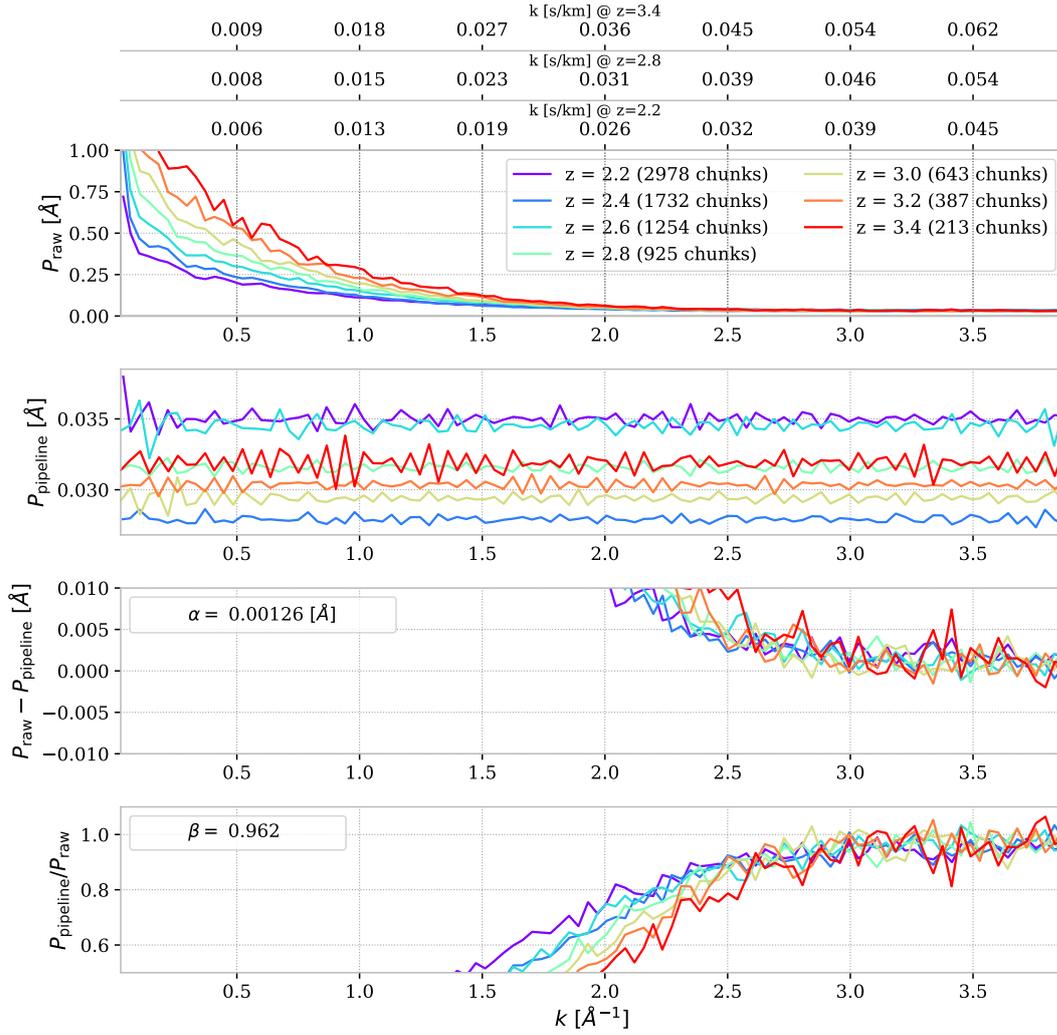


FIGURE 5.4 – Measurement of the asymptotic difference ( $\alpha$ ) and ratio ( $\beta$ ) between noise and raw power spectra. Here,  $\alpha$  and  $\beta$  are measured for the pipeline noise, using the SV3 observations with a  $\overline{\text{SNR}}$  cut of 3.

The exposure difference coadd is computed in `picca`<sup>5</sup> [14]. We obtain an estimator of  $P_{\text{noise}}$  called difference power spectrum and noted  $P_{\text{diff}}$ , such that:

$$P_{\text{diff}}(k) = \left| \mathcal{F} \left[ \frac{\Delta f_j^{\text{corr}}(\lambda)}{F(\lambda)C_q(\lambda)} - 1 \right] \right|^2. \quad (5.25)$$

### Comparison of noise estimate to high-wavenumber data

Fig. 5.3 shows the average correction due to resolution and pixelization for SV3 and DA0.2 data with  $\overline{\text{SNR}} > 3$ . This correction suppresses more than 95 % of the signal at  $k > k_{\text{res},95} = 2.69 \text{ \AA}^{-1}$ , and 98 % at  $k > k_{\text{res},98} = 3.11 \text{ \AA}^{-1}$ . This observation drives the maximal wavenumber that we will use for the measurement of  $P_{1D,\alpha}$ : here, we choose a value  $k_{\text{max}} = 2.5 \text{ \AA}^{-1}$ .

Additionally, the thermal broadening and Jeans smoothing suppresses the Ly $\alpha$  signal along the line-of-sight at small scales, as shown by simulation in Sec. 3.2.4.

We can take advantage of the combined effect of the resolution, pixelization, and thermal broadening to estimate the noise power spectrum level. Indeed, at large wavelength, Eqn. 5.13 simplifies into  $P_{\text{raw}}(k) \simeq P_{\text{noise}}(k)$ . We can use the ratio or difference of  $P_{\text{raw}}$  and the noise power estimator on the largest k-bins accessible to validate the noise estimator  $P_{\text{noise}}$ , and correct it empirically if needed. I define the following asymptotic values:

$$\begin{aligned}\alpha &= \langle P_{\text{raw}}(k) - P_{\text{noise}}(k) \rangle_{k > k_{\text{res},98}} , \\ \beta &= \left\langle \frac{P_{\text{noise}}(k)}{P_{\text{raw}}(k)} \right\rangle_{k > k_{\text{res},98}} .\end{aligned}\tag{5.26}$$

Fig. 5.4 shows the measurement of  $\alpha$  and  $\beta$  on SV3 data set with  $\overline{\text{SNR}} > 3$ , using the pipeline noise to compute  $P_{\text{noise}}$ . Considering the observed statistical fluctuations, we notice that in that case the asymptotic behavior of power spectra at high wavenumber enables a good measurement of  $\alpha$  (respectively  $\beta$ ), whose value is close to 0 (respectively 1). Moreover,  $\alpha$  and  $\beta$  do not vary as a function of redshift within statistical fluctuations.

As pointed out in Sec. 4.3.4.2, comparing the "difference" and "pipeline" noise power spectra was beneficial to improving the noise pipeline. With the Everest reduction, the correction of the master dark readout noise improved typical values of  $\beta$  from 0.5 to about 0.9. In the Fuji and Guadalupe reductions, considering the CCD position dependence of the master dark readout noise increased  $\beta$  values to more than 0.95.

The asymptotic quantities  $\alpha$  and  $\beta$  can be used to correct the missing noise not taken into account by the estimators. We correct this underestimation by adding in the estimation of  $P_{\text{1D},\alpha}$  (Eqn. 5.18) the following residual noise to  $P_{\text{noise}}$ :

$$P_{\text{noise,miss}} = \begin{cases} \alpha , & \text{additive correction ,} \\ \left( \frac{1-\beta}{\beta} \right) P_{\text{noise}} , & \text{multiplicative correction .} \end{cases}\tag{5.27}$$

### Residual noise correction

I computed the  $\alpha$  and  $\beta$  coefficients for the pipeline and difference noise estimators,  $P_{\text{pipeline}}$  and  $P_{\text{diff}}$ , on SV1, SV3, and DA0.2 data sets, with variation of the  $\overline{\text{SNR}}$  cut.

Fig. 5.5 shows fitted  $\alpha$  values for the Ly $\alpha$  forest band, using SV1, SV3 and DA0.2 with the  $P_{\text{pipeline}}$  and  $P_{\text{diff}}$  estimators. As previously stated,  $\alpha$  is equal to zero when the noise is correctly estimated. We can note that the values of  $\alpha$  are small compared to the absolute level of the noise power spectrum shown in Fig. 5.4. There exists a  $\overline{\text{SNR}}$  cut dependence for the SV1 data set, which is not present for DA0.2 and weakly present for SV3. This difference is probably related to the number of observations used to obtain the coadded spectra. We think that the source of unaccounted noise for SV1 originates from the use of a single master dark frame to estimate the dark current noise. Indeed, for this data set, the master dark is common to all exposures during spectra coaddition. Consequently, the source of noise produced by the dark current variations between exposures is not considered, and this effect increases for many exposures like SV1. The dependence in  $\overline{\text{SNR}}$  may be explained by the fact that this effect is amplified when we consider the noisier spectra. We choose to make a data set-dependent correction to remove this residual noise.

Still in Fig. 5.5, the power spectrum  $P_{\text{diff}}$  is represented for SV1 and SV3. In DA0.2, quasars are observed with one, or a low number of exposures ; thus,  $P_{\text{diff}}$  can not be reliably computed. For the SV1 and SV3 observations, the difference noise power spectrum exhibits the same trend as

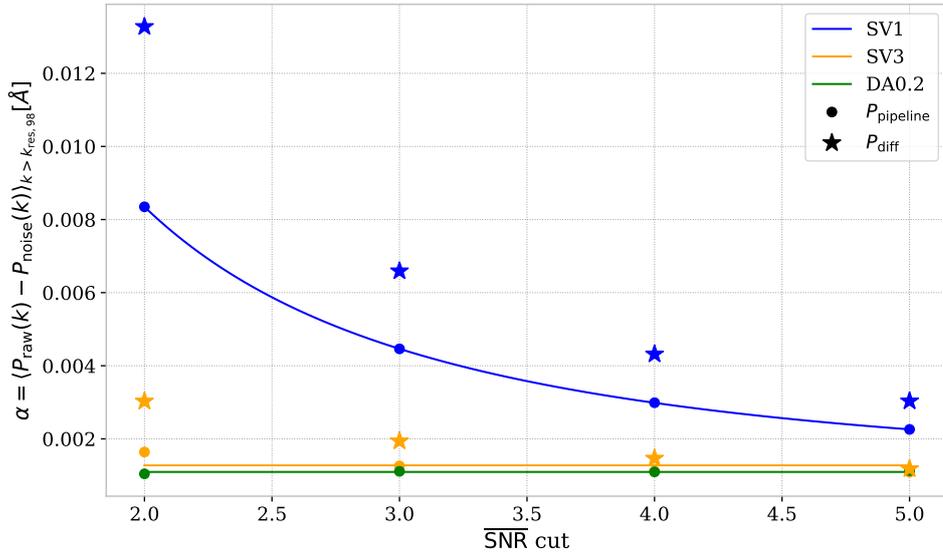


FIGURE 5.5 – Asymptotic differences  $\alpha$  between the noise and raw power spectra for SV1 (blue), SV3 (yellow), and DA0.2 (green) data sets, as a function of the  $\overline{\text{SNR}}$  cut. This difference is measured for both  $P_{\text{noise}}$  estimators from the pipeline (points) and from exposure differences (stars). The continuous lines are fits of the  $\alpha$  values for pipeline noise, whose parameters are given in Tab. 5.1.

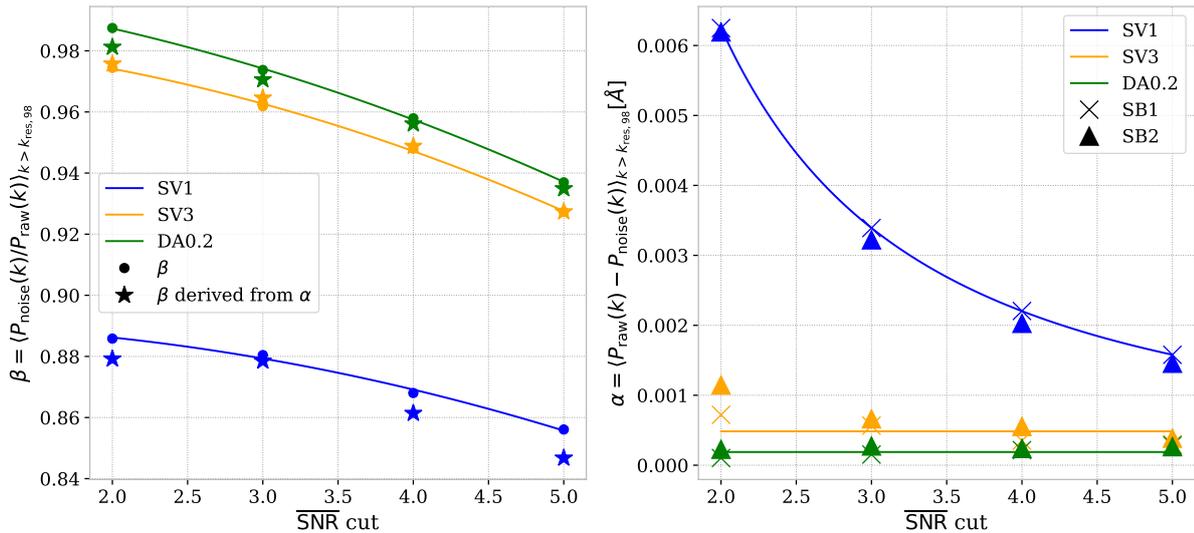


FIGURE 5.6 – (left) Asymptotic ratios  $\beta$  between the noise and raw power spectra for SV1 (blue), SV3 (yellow), and DA0.2 (green) data sets, as a function of the  $\overline{\text{SNR}}$  cut. Points give the direct estimation of  $\beta$ . Stars represent  $\beta$  as derived from the asymptotic differences in Fig. 5.5. Second-order polynomial fits are shown only for representation. (right) Asymptotic differences  $\alpha$  for SV1 (blue), SV3 (yellow), and DA0.2 (green) data sets in the side-band regions SB1 (crosses) and SB2 (triangles). The continuous lines are fits of the  $\alpha$  values, whose parameters are given in Tab. 5.1. For both panels, the pipeline noise is used.

the pipeline noise estimate. However,  $P_{\text{diff}}$  consistently underestimates the noise level compared to  $P_{\text{pipeline}}$ . As a consequence, we only use the  $P_{\text{pipeline}}$  estimation from now on.

Fig. 5.6 (left panel) shows the asymptotic ratios  $\beta$  for the three data sets. We remark that for SV3 and DA0.2, the  $\overline{\text{SNR}}$  dependence of  $\beta$  is much more pronounced than that of  $\alpha$ . Since

Band	Data	$P_{\text{noise,miss}} = \alpha [\text{\AA}]$
Ly $\alpha$	SV1	$0.026 \times (\overline{\text{SNR}})^{-1.77} + 0.00076$
	SV3	0.00127
	DA0.2	0.00109
Side-bands	SV1	$0.018 \times (\overline{\text{SNR}})^{-1.52} + 0.000032$
	SV3	0.00048
	DA0.2	0.00019

TABLE 5.1 – Additive corrections applied to the pipeline noise for different spectral regions and data sets. An  $\overline{\text{SNR}}$  dependence is included in the case of SV1 only. The same parameters are used for both SB1 and SB2.

the absolute noise level is the main parameter which varies when changing the  $\overline{\text{SNR}}$  cut, this is an indication that the residual noise source is additive rather than multiplicative. To support this hypothesis, we computed the  $\beta$  values derived from the  $\alpha$  of Fig. 5.5, using the mean value of  $P_{\text{noise}}$  for all redshift. They are shown as stars in the left panel of Fig. 5.6, and exhibit similar trends to the direct  $\beta$  computation, which corroborates that the missing noise is additive. Consequently, we decide to correct  $P_{\text{noise}}$  using an additive term  $\alpha$ .

I performed the same noise study in the rest-frame regions called side-bands SB1 ( $1270 < \lambda_{\text{rf}} < 1380 \text{ \AA}$ ), and SB2 ( $1410 < \lambda_{\text{rf}} < 1520 \text{ \AA}$ ), illustrated in Fig. 2.3, and which are used in the estimation of the metal power spectrum in Sec. 5.2.2. The  $\alpha$  values are shown in the right panel of Fig. 5.6. For side-bands, the overall missing noise level is lower than for the Ly $\alpha$  band, but exhibits similar trends as a function of  $\overline{\text{SNR}}$  cut. This difference could be due to the use of rather different quasar redshift populations between side-bands and the Ly $\alpha$  forest. As the results for SB1 and SB2 are very similar, we decide to apply the same correction for these two bands.

Tab. 5.1 summarizes the corrections to the pipeline noise we computed for the different data sets. For SV3 and DA0.2, a constant additive correction ( $\alpha$ ) is used. For SV1, given the observed dependence of  $\alpha$  as a function of  $\overline{\text{SNR}}$  cut, we choose to fit it with a power-law.

## 5.2.2 Side-band power spectrum

Following previous studies [2, 3, 4], we use side-bands to estimate the power spectrum components  $P_{\text{metals}}$  due to the absorption of metals in the Ly $\alpha$  forest. We define the **side-bands** SB1 ( $1270 < \lambda_{\text{rf}} < 1380 \text{ \AA}$ ) and SB2 ( $1410 < \lambda_{\text{rf}} < 1520 \text{ \AA}$ ) as shown on Fig. 2.3. In both side-bands, the fraction of transmitted flux contrast can be expressed similarly to Eqn. 5.11:

$$\delta_F(\lambda)|_{\text{SB}} = \delta_{\text{metals}}(\lambda)|_{\text{SB}} \otimes W(\lambda, \mathbf{R}) \otimes \Pi(\lambda, \Delta\lambda_{\text{pix}}) + \delta_{\text{noise}}(\lambda)|_{\text{SB}}. \quad (5.28)$$

The  $\delta_{\text{metals}}(\lambda)|_{\text{SB}}$  contrast contains all the fluctuations due to metals with rest-frame absorption wavelength higher than  $1380 \text{ \AA}$  for SB1, and  $1520 \text{ \AA}$  for SB2. Similarly to the calculation of  $P_{1D,\alpha}$ , the side-band power spectrum then writes:

$$P_{\text{SB}}(k) = \left\langle [P_{\text{raw}}(k)|_{\text{SB}} - P_{\text{noise}}(k)|_{\text{SB}}] \cdot \mathbf{R}^{-2}(k) \cdot \text{sinc}^2\left(\frac{k\Delta\lambda_{\text{pix}}}{2}\right) \right\rangle. \quad (5.29)$$

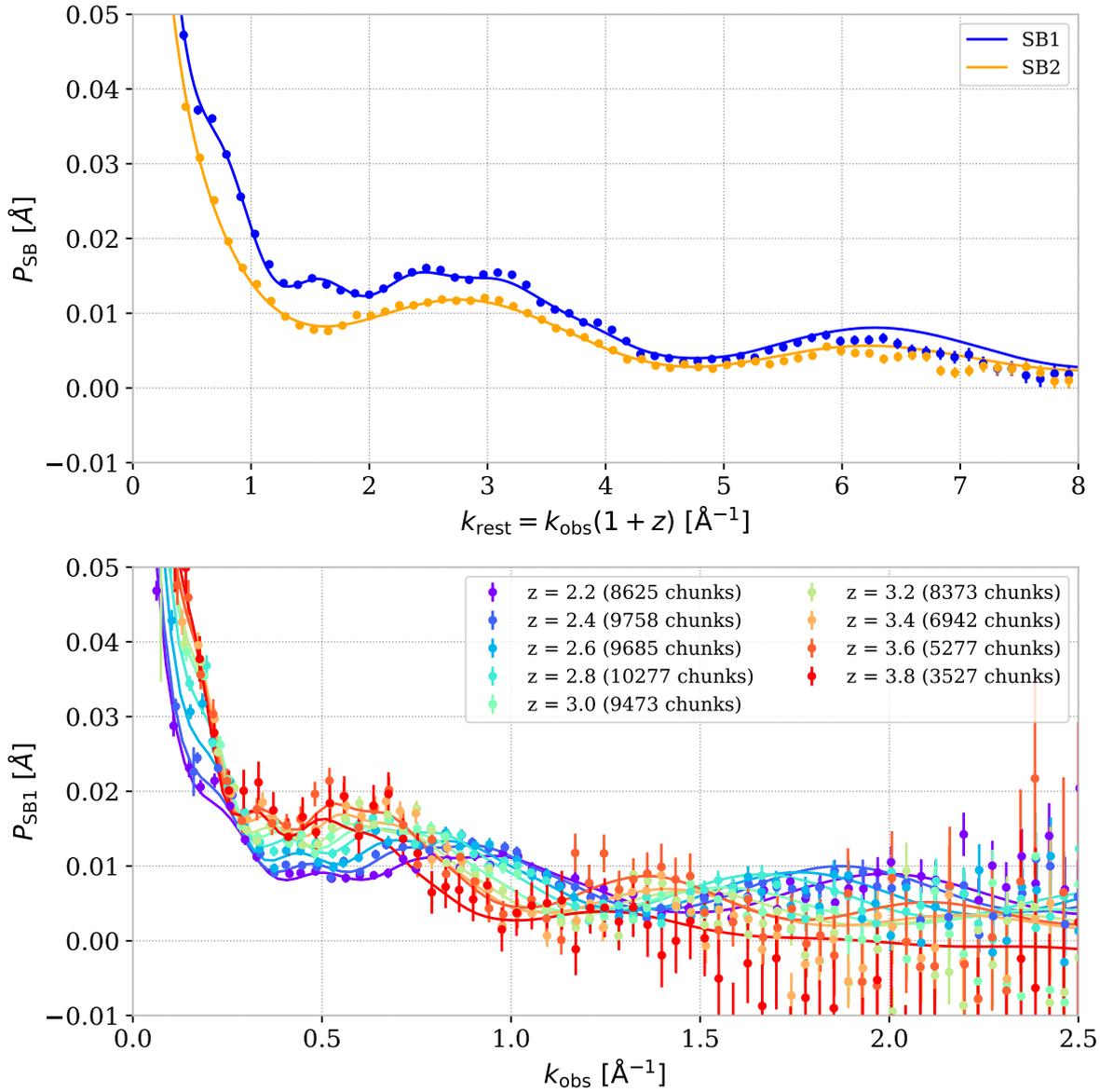


FIGURE 5.7 – One-dimensional power spectra measured in the side-band regions SB1 and SB2, using the combination of SV3 and DA0.2 data sets after applying the dataset-dependent noise correction. The SNR cut is redshift dependent, following Tab. 5.2. (top) Average of  $P_{\text{SB}}$  over all redshift bins in the rest wavenumber frame  $k_{\text{rest},i} = k_{\text{obs},i} \times (1 + z)$ . The fitted model represented in continuous line is given by Eqn. 5.30. (bottom)  $P_{\text{SB}}$  on side-band SB1 as a function of redshift and observed wavenumber. Each redshift bin is fitted using the product between Eqn. 5.30, with parameters fixed from previous fit, and a first-order polynomial function.

The SB1 power spectrum includes SiIV absorptions, one of the major metals, which is not present in SB2. Consequently, we use the side-band power spectrum of SB1 to estimate  $P_{\text{metals}}$  in Eqn. 5.18, and the SB2 power spectrum as a consistency check. The quasars used to calculate this side-band power spectrum are not located at the same redshift as those used to compute  $P_{\text{1D},\alpha}$ . In particular, quasars at  $z < 2.0$  are used to calculate  $P_{\text{SB}}$  in the lowest redshift range of  $P_{\text{1D},\alpha}$ . However, the intergalactic medium which produces the metal power spectrum is at

the same redshift for  $P_{\text{SB}}$  and  $P_{\text{ID},\alpha}$  calculation, which implies that the redshift dependence of metal absorptions is correctly taken into account.

### Measurement and modeling of side-bands

We present here the measurement of the side-band power spectrum using SV3 and DA0.2 data (see Sec. 5.4), with redshift dependent  $\overline{\text{SNR}}$  cuts given in Tab. 5.2, and a noise estimated by the pipeline and corrected as given by Tab. 5.1. The result is shown in Fig. 5.7. The top panel shows the stack by interpolation of  $P_{\text{SB}}$  for all redshift bins in wavenumber rest-frame space  $k_{\text{rest}} = (1+z) \times k_{\text{obs}}$ . The bottom panel shows separated redshift bins as a function of observational wavenumber. A total number of 93,127 and 134,153 sub-forests were used for SB1 and SB2 measurements, respectively. In the Fig. 5.7 top panel, the average side-band power is lower for the SB2 than SB1, as expected by the addition of SiIV absorptions.

In the eBOSS measurement [3], a sixth-degree polynomial is used to fit the shape of the side-band power. For our measurement, we use the stacked  $P_{\text{SB}}$  profile to design a more physically motivated model.

The complete list of metals that impact  $P_{\text{ID},\alpha}$  is given in [17] and the strongest absorptions are from SiIII, SiII, SiIV, and CIV. The emission peaks, and consequently absorption peaks, of SiIV and CIV are actually two doublets. Their rest-frame wavelength given by NIST [18] are  $\lambda_{\text{SiIV}^a} = 1393.76 \text{ \AA}$ ,  $\lambda_{\text{SiIV}^b} = 1402.77 \text{ \AA}$ ,  $\lambda_{\text{CIV}^a} = 1548.202 \text{ \AA}$ , and  $\lambda_{\text{CIV}^b} = 1550.774 \text{ \AA}$ . The presence of an absorption doublet in the side-band creates a peak in the one-dimensional correlation function, which translates into an oscillatory patterns in the power spectrum, whose periodicity depends on the doublet separation. In the top panel of Fig. 5.7, both side-band power spectra display a large oscillation caused by the CIV doublet. As expected, the SB1 power spectrum shows an additional oscillation induced by SiIV doublet absorptions.

These considerations lead us to model  $P_{\text{SB}}$  as the sum of a power-law including all-metal contributions and oscillations due to SiIV and CIV doublets:

$$P_{\text{SB},m} = A \times k^{-\epsilon} + \sum_i P_{\text{doublet},i}(k, A_i, a_i, k_i, \psi_i). \quad (5.30)$$

Oscillations induced by a doublet have a frequency characterized by the rest-frame wavenumber  $k_{\text{rest},i} = 2\pi/\delta\lambda_i$  where  $\delta\lambda_i$  is the doublet wavelength difference in  $\text{\AA}$ . We choose to use damped sinusoidal functions to model the doublet oscillations as follow:

$$P_{\text{doublet},i}(k, A_i, a_i, k_i, \psi_i) = A_i e^{-a_i k} \sin\left(2\pi\left(\frac{k}{k_i}\right) + \psi_i\right), \quad (5.31)$$

where  $k_i$  is a free parameter with a uniform prior centered around  $k_{\text{rest},i}$ .

The result of the fit on the redshift-averaged  $P_{\text{SB}}$ , taking into account the oscillations of CIV and SiIV for SB1, and only CIV for SB2, is shown in the top panel of Fig. 5.7. The SB1 fitted function is used to derive the redshift dependence of the side-band power spectrum, shown in the bottom panel of Fig. 5.7. For each redshift bin, we fit a product between the global SB1 fitted function (expressed in the observational wavelength frame), and a first-order polynomial. By choosing to have so many fit parameters, and in particular, allowing  $k_i$  to vary, we somewhat overfit the data: this is on purpose, as  $P_{\text{SB}}$  may also include other uncorrelated contaminations and systematics besides metals. Our fit is designed to closely reproduce the  $P_{\text{SB}}$  data points so that we can use it as a statistical subtraction for the measurement of  $P_{\text{ID},\alpha}$  in Eqn. 5.18.

Our  $P_{\text{SB}}$  measurement already represents a clear improvement with respect to that of BOSS [2], and eBOSS [3]: by eye, SiIV and CIV induced oscillatory patterns are seen even

for individual redshift bins, and  $P_{\text{SB}}$  is essentially a decreasing function of wavenumber, even at high wavenumber. This indicates an improvement in the noise modeling.

### 5.2.3 Sky emission lines

Atmospheric sky emission lines are removed from DESI spectra by the spectral extraction pipeline as described in Sec. 4.3.3. This is achieved by observing sky spectra with dedicated fibers pointing at regions with no background sources. The average of 15,000 sky spectra on exposures with optimal observing conditions, noted  $\langle f_{\text{sky}} \rangle$ , is shown in Fig. 5.8.

The noise of pixels associated to intense sky lines is strongly increased and impacts the  $P_{1\text{D},\alpha}$  calculation. We choose to mask the major sky lines as in previous measurements [2, 3]. The catalog of lines used in these studies<sup>7</sup> was adapted to the spectral resolution of the SDSS instrument. The improved resolution of DESI makes it possible to reduce the masking size for narrow sky lines. Reducing this size decreases the impact of masking on the  $P_{1\text{D},\alpha}$  measurement.

#### Sky line finder

I developed an algorithm similar to [19] to compute a sky line catalog adapted to the DESI instrument. I apply a median smoothing  $\mathcal{M}(d_\lambda)$ , of spectral width  $d_\lambda = 160 \text{ \AA}$ , on the average sky spectrum. I select a sky line when the average sky spectrum is larger than the product of the

7. [https://github.com/igmhub/picca/blob/master/etc/list\\_veto\\_line\\_Pk1D.txt](https://github.com/igmhub/picca/blob/master/etc/list_veto_line_Pk1D.txt)

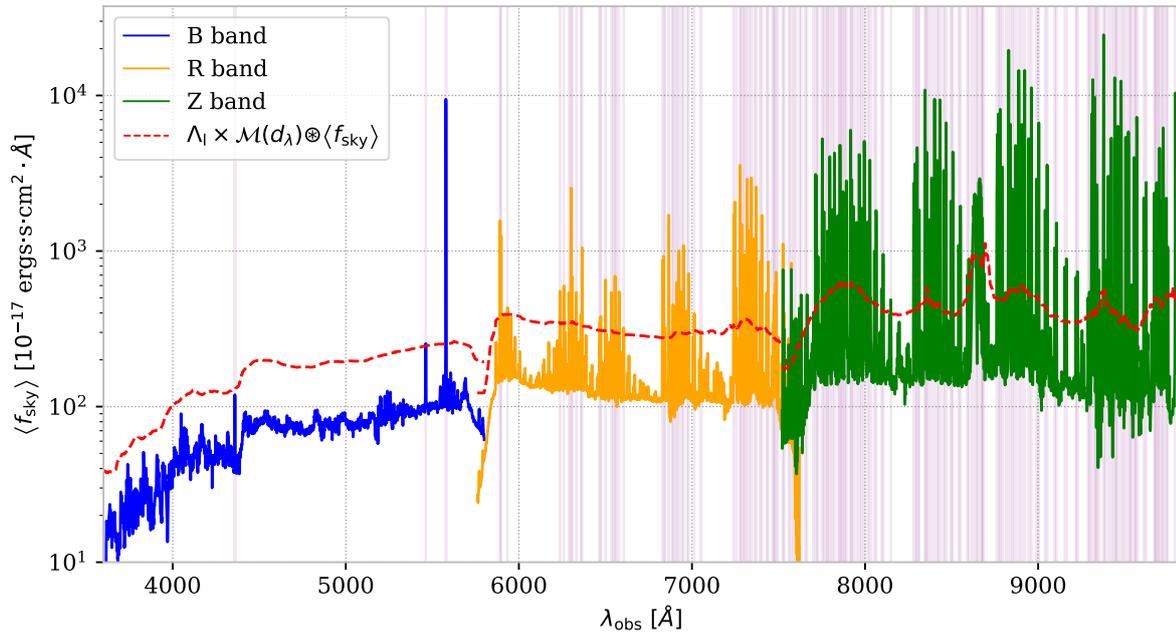


FIGURE 5.8 – Average of the sky spectra of 15,000 sky fibers with optimal observation conditions (speed  $> 105$ , exposure time  $> 1100$  sec, seeing  $< 1.05^\circ$ , and airmass  $< 1.3$ ). These sky spectra originate from three exposures in the SV1 and SV3 data sets with Fuji pipeline reduction (EXPID = 84235, 87116, 87600). The different spectral bands of DESI are represented (B in blue, R in orange, and Z in green). The median smoothing of this average sky spectrum multiplied by a threshold  $\Lambda_1 = 2.5$ , shown as a dashed red line, is used to select sky emission lines. The line located at  $4360 \text{ \AA}$  was added manually, considering its large impact on noise seen in Fig. 5.9.

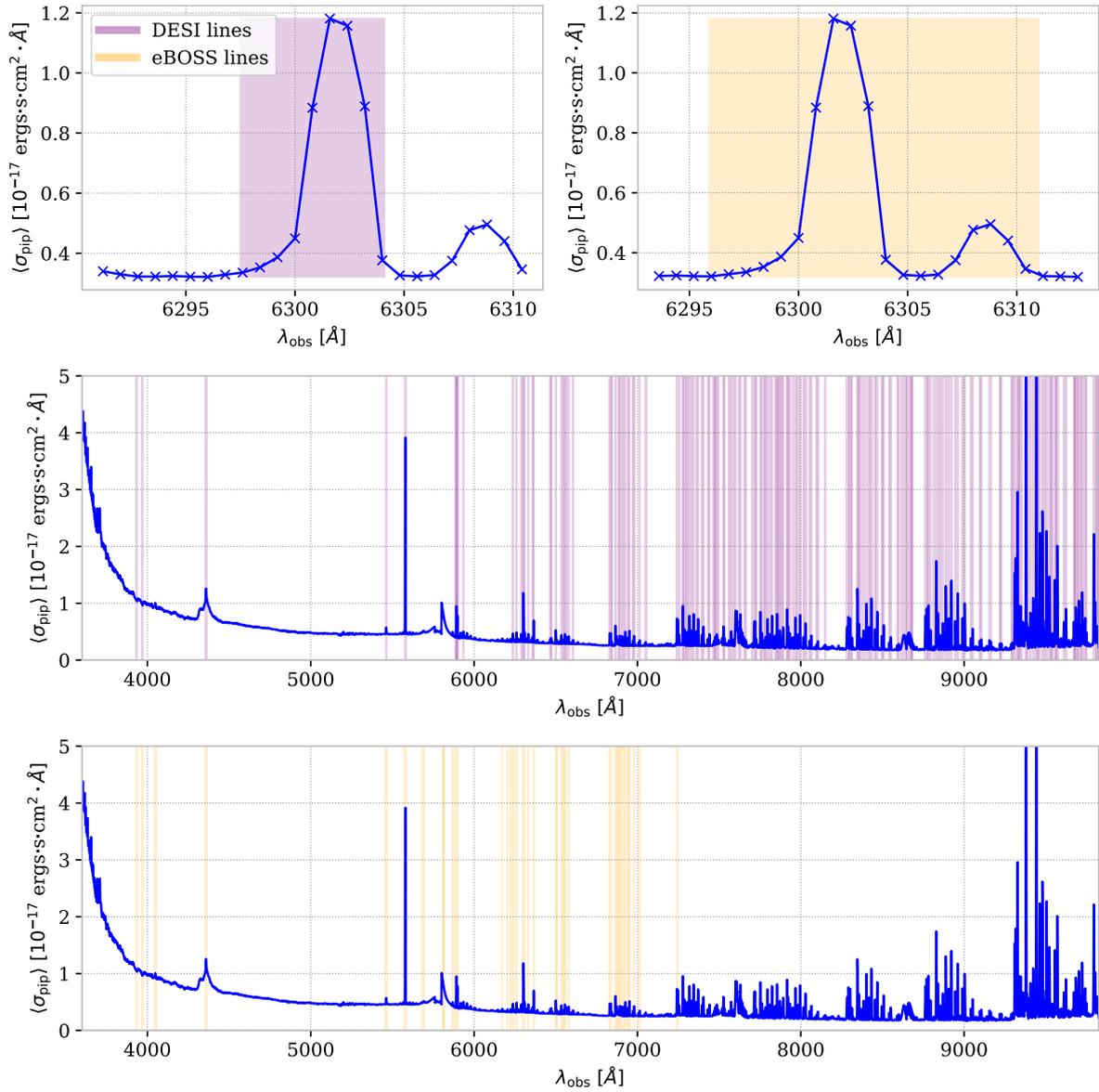


FIGURE 5.9 – Comparison between the sky emission line catalogs used for SDSS<sup>7</sup> (orange) and the new DESI catalog we designed here (purple). The blue curves represent the DESI average pipeline noise of 125,477 objects categorized by **redrock** as QSOs or LRGs on SV3 observations. (top) Zoom on a specific sky emission line for which the DESI catalog minimizes the masked length. (middle) DESI ( $\Lambda_1 = 2.5$ ,  $\Lambda_w = 1.2$ ) line catalog. (bottom) SDSS<sup>7</sup> line catalog.

smoothed sky flux,  $\mathcal{M}(d_\lambda) \otimes \langle f_{\text{sky}} \rangle$ , by a threshold  $\Lambda_1$ . In Fig. 5.8, the dashed red line represents this product for  $\Lambda_1 = 2.5$  and  $d_\lambda = 160 \text{ \AA}$ .

A second threshold  $\Lambda_w = 1.2$  is used to define the width of sky lines. The upper and lower wavelength limits of a sky line are defined as the first wavelengths on each side whose average sky spectrum is lower than  $\Lambda_w \times \mathcal{M}(d_\lambda) \otimes \langle f_{\text{sky}} \rangle$ . To remain conservative and prevent numerical effects potentially caused by masking at a pixel position, I choose to increase the width of all lines by  $1 \text{ \AA}$  on each side.

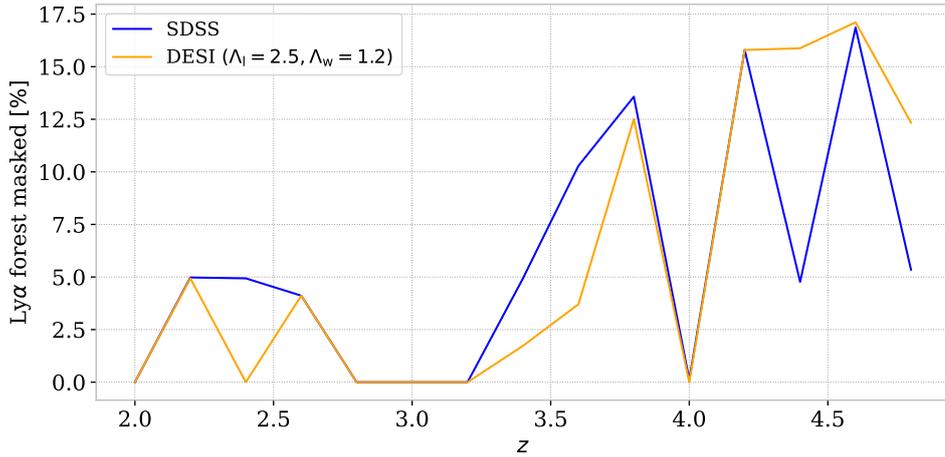


FIGURE 5.10 – Percentage of spectral length masked by sky lines as a function of redshift, for SDSS<sup>7</sup> and DESI ( $\Lambda_1 = 2.5, \Lambda_w = 1.2$ )<sup>8</sup> catalogs.

In this sky line catalog, I also add the galactic absorption lines, which correspond to relatively broad absorptions made by dust in the Milky Way. We take the same lines as SDSS<sup>7</sup>: CaII H and K lines at 3968 and 3933 Å, and the NaD doublet at 5893 Å.

Finally, the DESI sky emission line catalog built from this procedure is available online<sup>8</sup>.

### Visual inspection

I checked that the produced sky line catalog correctly masks the DESI noise. For this purpose, I compute an average noise by stacking the pipeline noise  $\sigma_{\text{pip}}$  of 125,477 objects categorized by `redrock` as QSOs or LRGs on SV3 observations. This average noise with the DESI sky line catalog<sup>8</sup> such that ( $\Lambda_1 = 2.5, \Lambda_w = 1.2$ ) is shown in the middle panel of Fig. 5.9. For comparison, the SDSS catalog<sup>7</sup> is represented in the bottom panel.

As an example, a zoom on a specific sky line is shown in the top panels of Fig. 5.9 with SDSS and DESI masks. The SDSS mask is too wide for DESI, which highlights the benefit of creating a new catalog. After a visual inspection of most sky lines, I validate that all the pixels showing an increase in noise are masked by setting the width threshold to  $\Lambda_w = 1.2$ .

Comparing the average noise (Fig. 5.9) to the average sky flux (Fig. 5.8), there is a consistency between sky emission lines and observed peaks in the pipeline noise. The feature at 4360 Å is an exception, as it appears wide and relatively high in the DESI noise and not in the average sky flux. Its wavelength is inside a well-known transmission dip around 4400 Å due to an issue with DESI’s spectrograph collimator coating. For this specific line, we take the same value as the SDSS catalog and force the algorithm to consider it as a sky line even if it does not pass the  $\Lambda_1$  requirement.

A comparison of the percentage of line-of-sight masked for SDSS and DESI catalog, as a function of redshift, is given in Fig. 5.10.

8. [https://github.com/corentinravoux/pldesi/blob/main/etc/skylines/list\\_mask\\_Pk1D\\_DESI\\_SV\\_fuji\\_threshold\\_2.5\\_2.5\\_2.5\\_medianfilter200\\_threshold\\_width\\_1.2\\_1.2\\_1.2.txt](https://github.com/corentinravoux/pldesi/blob/main/etc/skylines/list_mask_Pk1D_DESI_SV_fuji_threshold_2.5_2.5_2.5_medianfilter200_threshold_width_1.2_1.2_1.2.txt)

## 5.3 Characterization on mocks

The results presented in this section were obtained by Marie-Lynn Abdul Karim, an M2 student whom I contributed to supervise during 3 months in Spring 2022.

We use mocks specifically designed for  $P_{1D,\alpha}$ , described in Sec. 3.1.2. In particular, we use the Y5 realization of **DESI-Lite**, which mimics the statistics and redshift distribution of Ly $\alpha$  forest data at the end of the DESI survey. For all the results on the mocks presented here, we performed a  $\overline{\text{SNR}}$  cut of 3 for all redshifts, which is the average cut imposed on data. We checked that the corrections developed in this section do not depend on this  $\overline{\text{SNR}}$  cut. We performed a linear rebin to obtain a wavenumber binning three times coarser than the data.

The mocks are used to characterize the impact of pixel masking (Sec. 5.3.1) and residual effects such as continuum fitting impact (Sec. 5.3.2) on the  $P_{1D,\alpha}$ , and compute empirical corrections of these effects.

### 5.3.1 Pixel masking

For both HCD and sky line masking, we remove some data points from the measured spectra. This is harmless for likelihood inference such as continuum fitting or the quadratic maximum likelihood estimator for  $P_{1D,\alpha}$  [5]. On the other hand, the FFT calculation requires that Ly $\alpha$  contrast pixels are equally-spaced. Consequently, when computing Fourier transform, we impose a null Ly $\alpha$  contrast and infinite standard deviation to the masked pixels. This masking introduces a  $k$ -dependent bias which we need to quantify.

In order to determine and correct this bias in our  $P_{1D,\alpha}$  measurement, we created a mock that includes only noise and the quasar's continuum modeled by **quickquasars**. By additionally imposing the masking on these mocks, it is possible to measure the masked power spectrum  $P_{1D,\alpha,\text{masked}}$ . The coefficient used for the masking correction is defined as the ratio between this power spectrum and the unmasked power spectrum  $P_{1D,\alpha,\text{unmasked}}$ :

$$A_{\text{mask}}(k, z) = \frac{P_{1D,\alpha,\text{unmasked}}(k, z)}{P_{1D,\alpha,\text{masked}}(k, z)}. \quad (5.32)$$

#### Sky emission lines

The correction induced by the DESI sky line mask, as defined in Sec. 5.2.3, is shown in Fig. 5.11. We verified that, for all redshifts where the masked line-of-sight length of DESI is close to SDSS, the impact of masking is lower in the DESI case. It indicates that obtaining thinner lines mitigates the impact of masking.

As expected, the correction roughly scales with the number of masked pixels. The effect of masking is a relatively smooth function of wavenumber, and its main impact is at low wavenumber. The most impacted redshift bins are  $z = 2.2$  (CaII galactic absorption lines),  $z = 2.6$  (lines at 4360 Å in the transmission dip), and at high redshift for which many sky lines need to be masked. The impact of sky line masking is qualitatively in agreement with the SDSS results in [3]. We choose to model  $A_{\text{sky}}(k, z)$  by a second-order polynomial fit and use this correction in the final calculation of  $P_{1D,\alpha}$ .

#### HCD masking

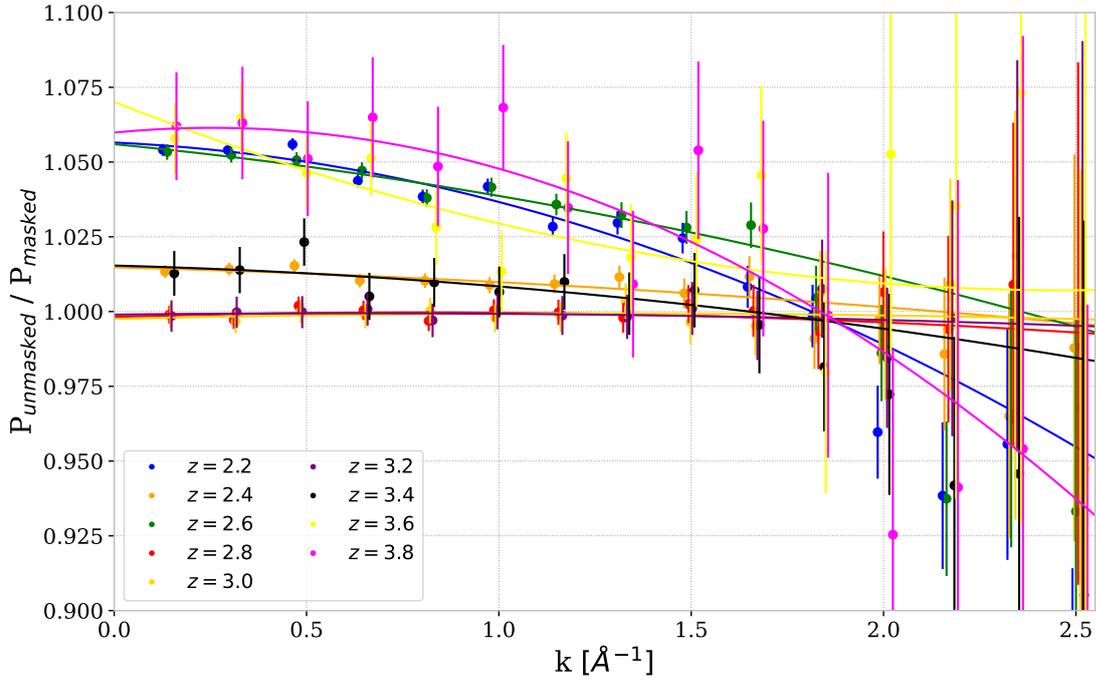


FIGURE 5.11 – Ratio between the unmasked and masked power spectra for sky line masking on the Y5 mocks. The DESI ( $\Lambda_l = 2.5, \Lambda_w = 1.2$ )<sup>8</sup> sky line catalog is used and a  $\overline{\text{SNR}}$  cut of 3 is imposed. Each power spectrum is re-binned by a factor 3 to reduce error bars. Second-order polynomial functions are used to fit the corrections in each redshift bin.

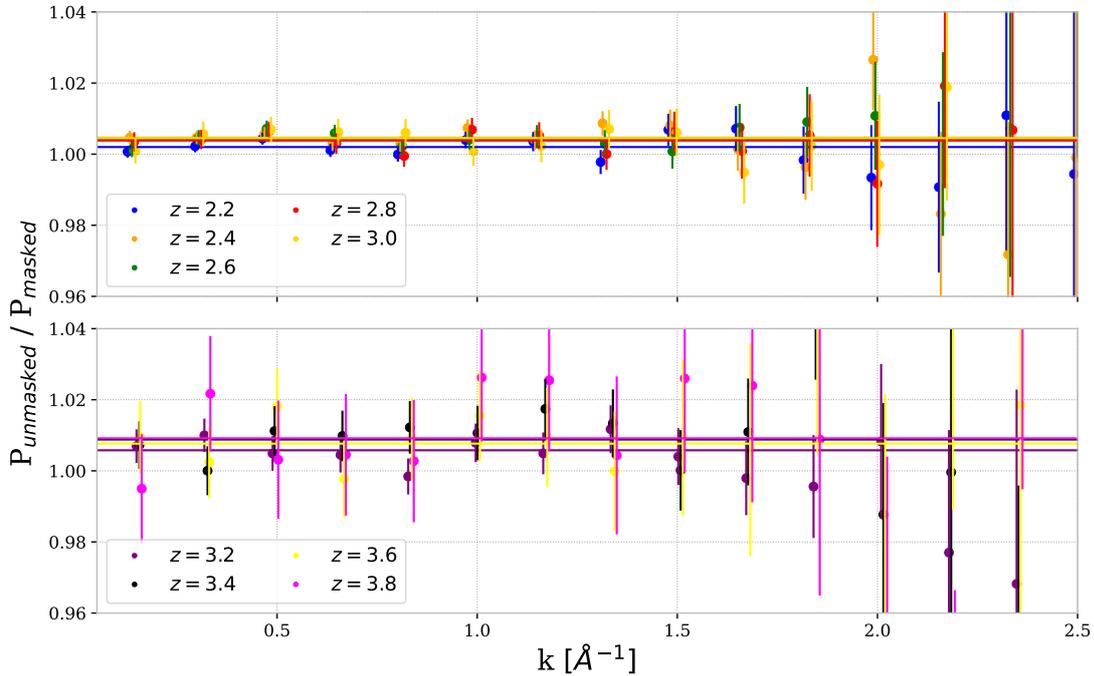


FIGURE 5.12 – Ratio between the unmasked and masked power spectra for HCD masking on the Y5 mocks. These ratios are computed using the same methodology as in Fig. 5.11. Continuous lines shows constant fits used for correction.

HCD objects are added at random locations in the Ly $\alpha$  forest during the creation of the mocks (see Sec. 3.1.2). Since we want to characterize the impact of HCD masking only, we apply the masking to a mock sample for which no HCD were added. For the moment, we do not attempt to characterize the completeness of the HCD finder used for the data, and we use a "truth" HCD catalog, built following the distribution in redshift and column density measured in eBOSS [12].

We mask the "truth" catalog with the same parameters as the HCD data catalog, for  $N_{\text{HI}} > 10^{20.3} \text{ cm}^{-2}$ . The correction induced by the masking,  $A_{\text{hcd}}(k, z)$ , is represented in Fig. 5.12. As it was already seen in the eBOSS measurement [3], the HCD masking has a smaller impact than sky emission lines. This is due to the random distribution of HCD and the smaller masking in terms of Ly $\alpha$  forest length. Given the smallness of the correction, and its statistical uncertainty, we apply to  $P_{\text{1D},\alpha}$  a  $k$ -independent correction  $A_{\text{hcd}}(k, z) = A_{\text{hcd}}(z)$ , whose amplitude is 0.5 % on average.

Note that the performance of the DLA finder in terms of completeness has not been studied here.

### 5.3.2 Continuum fitting

The continuum fitting procedure defined in Sec. 5.1.2 may bias  $P_{\text{1D},\alpha}$  because noise in the data together with the fitting procedure may systematically distort the measured continuum  $C_q(\lambda, z_q)\bar{F}(\lambda)$ . This is a well-known effect in BAO measurements [13].

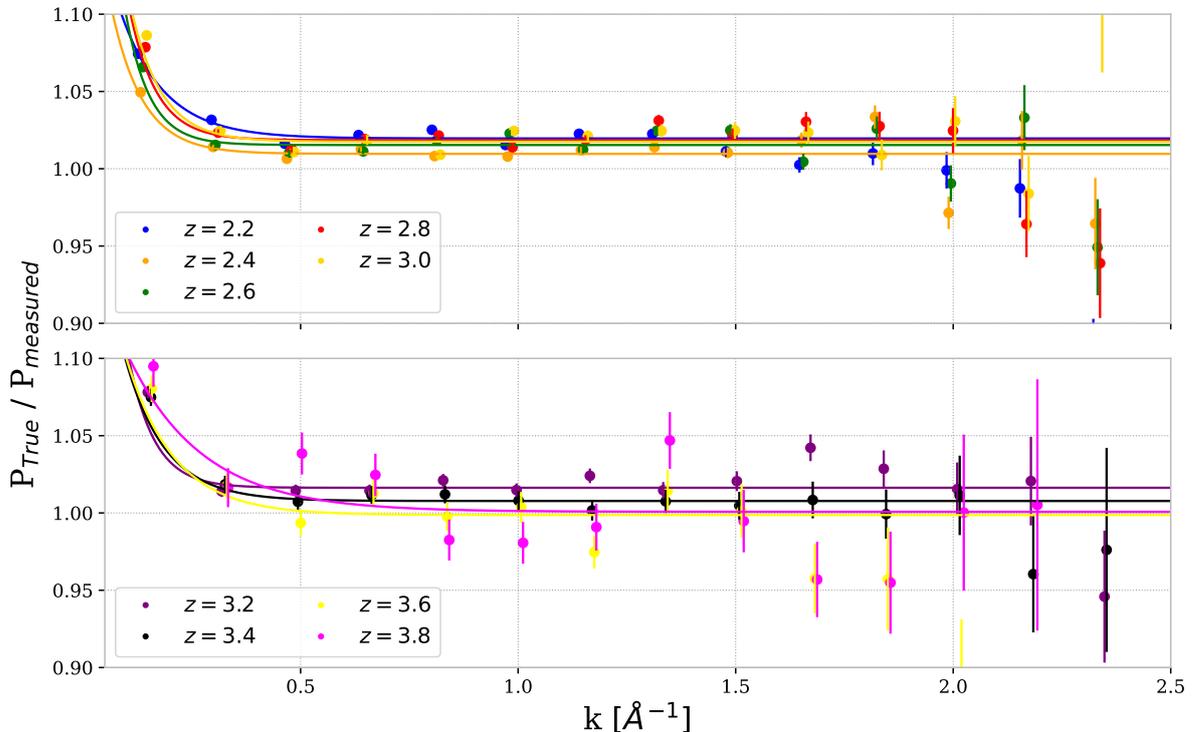


FIGURE 5.13 – Ratios between the true input power spectrum and the one derived with our pipeline, on the Y5 mocks. These ratios are computed using the same methodology as for pixel masking. Fitting functions are represented by continuous lines, and used to correct the  $P_{\text{1D},\alpha}$  measurement.

The resolution correction modeling, as well as the use of the FFT method could also potentially bias the  $P_{1D,\alpha}$  measurement. In this work, we do not seek to decouple the distinctive effects related to FFT method, resolution, and continuum fitting. We simultaneously take into account all of them by comparing the true input power spectrum  $P_{1D,\alpha,\text{true}}(k, z)$ , presented in Sec. 3.1.2 for DESI-Lite mocks, to the associated  $P_{1D,\alpha}$  measurement, done with the standard continuum fitting procedure  $P_{1D,\alpha,\text{measured}}(k, z)$ . We compute the corrective coefficient

$$A_{\text{residual}}(z, k) = \frac{P_{1D,\alpha,\text{true}}(k, z)}{P_{1D,\alpha,\text{measured}}(k, z)}. \quad (5.33)$$

This corrective factor  $A_{\text{residual}}(z, k)$  is shown in Fig. 5.13. We use an exponential function with the shape  $p_0 \exp(-k/p_1) + p_2$  to fit this correction and apply it to the  $P_{1D,\alpha}$  measurement.

Contrarily to SDSS measurement [3], the true input power spectrum is always larger than the measured one. The origin of this discrepancy is under investigation. Moreover, decoupling the effects of the FFT method, the resolution, and the continuum fitting is a work in progress.

## 5.4 One-dimensional Lyman- $\alpha$ power spectrum measurement on DESI data

SV1, SV3, and DA0.2 are very different data sets, as described in Sec. 4.3.1: they have different target selections and exposure times and were collected for different states of the DESI instrument. From a  $P_{1D,\alpha}$  point of view, the noise properties of SV1 are a potential issue. While our initial goal was to analyze the combined SV+DA0.2 data sample, we choose first to compare the measured  $P_{1D,\alpha}$  on the separate SV1, SV3, and DA0.2 data sets. Fig. 5.14 shows their respective ratios, on the four redshift bins with largest statistics. It appears that the measurement of  $P_{1D,\alpha}$  on SV1 is biased compared to the other two data sets. In particular, we believe that the difference at  $k \gtrsim 1.0 \text{ \AA}^{-1}$  is due to an imperfection in the noise correction presented in Sec. 5.2.1. Consequently, we decide to remove the SV1 data set in this study to remain conservative.

This section presents the application of the methodology and corrections described in previous sections to the combination of SV3, and DA0.2 data sets, noted SV3+DA0.2.

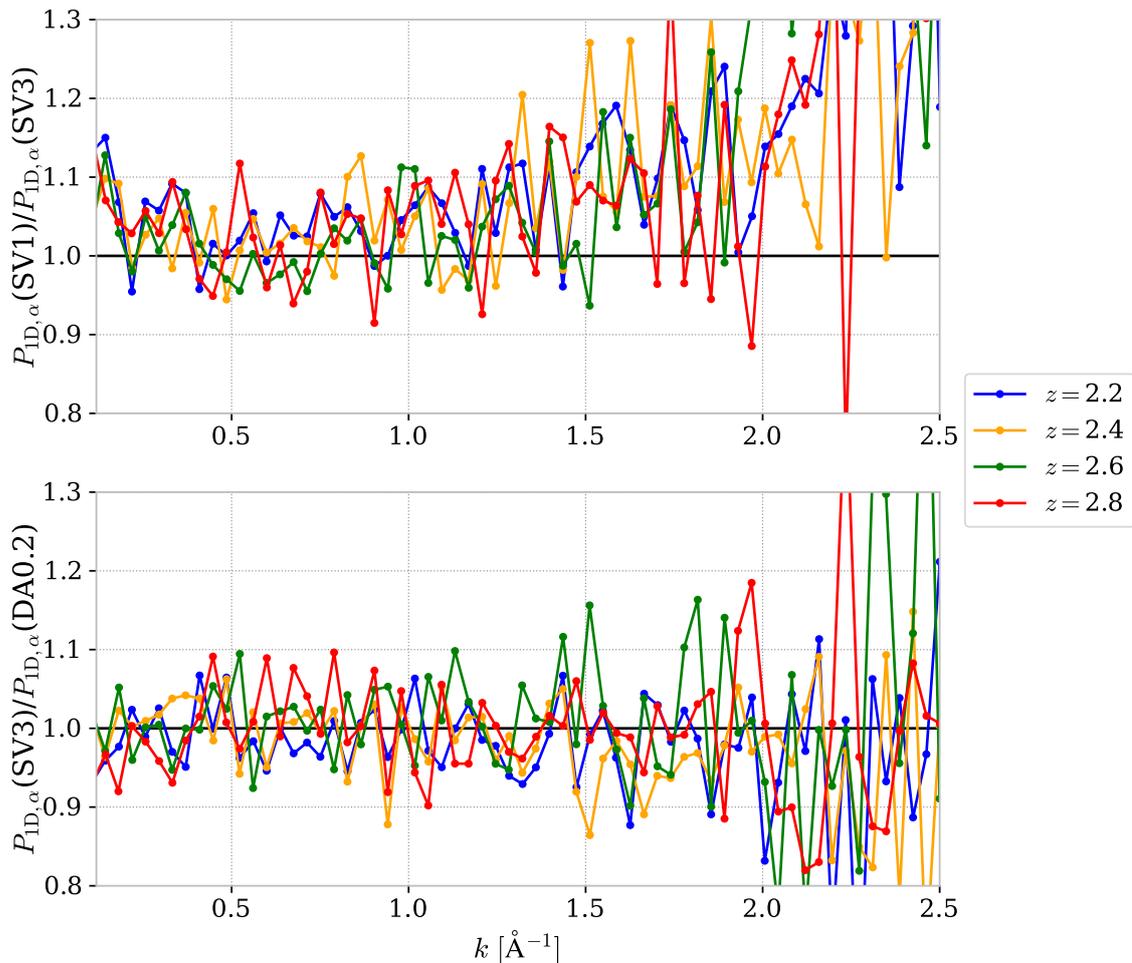


FIGURE 5.14 – Ratio of the  $P_{1D,\alpha}$ , measured with the same parameters, between SV1 and SV3 (top), and SV3 and DA0.2 (bottom), for four redshift bins.

## 5.4.1 Determination of uncertainties

### Statistical uncertainties

For each  $(k, z)$  bin, the statistical uncertainty of our  $P_{1D,\alpha}$  measurement, that we note  $\sigma_{\text{stat}}$ , is computed from the rms of the distribution of  $P_{1D,\alpha}$  over all forests. Fig. 5.15 shows  $\sigma_{\text{stat}}$ , computed as a function of the  $\overline{\text{SNR}}$  threshold (defined by Eqn. 5.8) for the small and large scales. We find similar trends to the latest SDSS measurement [3]: for large scales ( $k < 0.6 \text{ \AA}^{-1}$ ) where  $P_{\text{raw}}$  dominates over  $P_{\text{noise}}$ , the statistical error tends to increase with the  $\overline{\text{SNR}}$  threshold. Conversely, the noise dominates for small scales ( $k > 1.8 \text{ \AA}^{-1}$ ), and increasing  $\overline{\text{SNR}}$  cuts tends to decrease error bars. The latter is not true for high redshift, for which the statistics are low. In this case, increasing  $\overline{\text{SNR}}$  cut drastically reduces the number of Ly $\alpha$  forests and increases the standard deviation. We choose to apply the same  $\overline{\text{SNR}}$  thresholds as in [3], reported in Tab. 5.2. We verified that they minimize the statistical uncertainties for a central mode  $k = 1.2 \text{ \AA}^{-1}$ .

After applying this redshift-dependent  $\overline{\text{SNR}}$  cut, the statistical uncertainties associated with the  $P_{1D,\alpha}$  measurement are given in Fig. 5.16. The statistical error depends mainly on the number of chunks for each redshift bin, so that  $\sigma_{\text{stat}}$  is an increasing function of redshift. An exception is the case of  $z \sim 2.2 - 2.4$  at  $k \gtrsim 1.5 \text{ \AA}^{-1}$ : this is caused by the large noise increase in the blue spectral band due to atmospheric absorptions.

Looking at wavenumber dependence,  $\sigma_{\text{stat}}$  increases as a function of  $k$  for small scales ( $k \gtrsim 2 \text{ \AA}^{-1}$ ). This is due to the resolution correction, which effectively increases the rms of individual

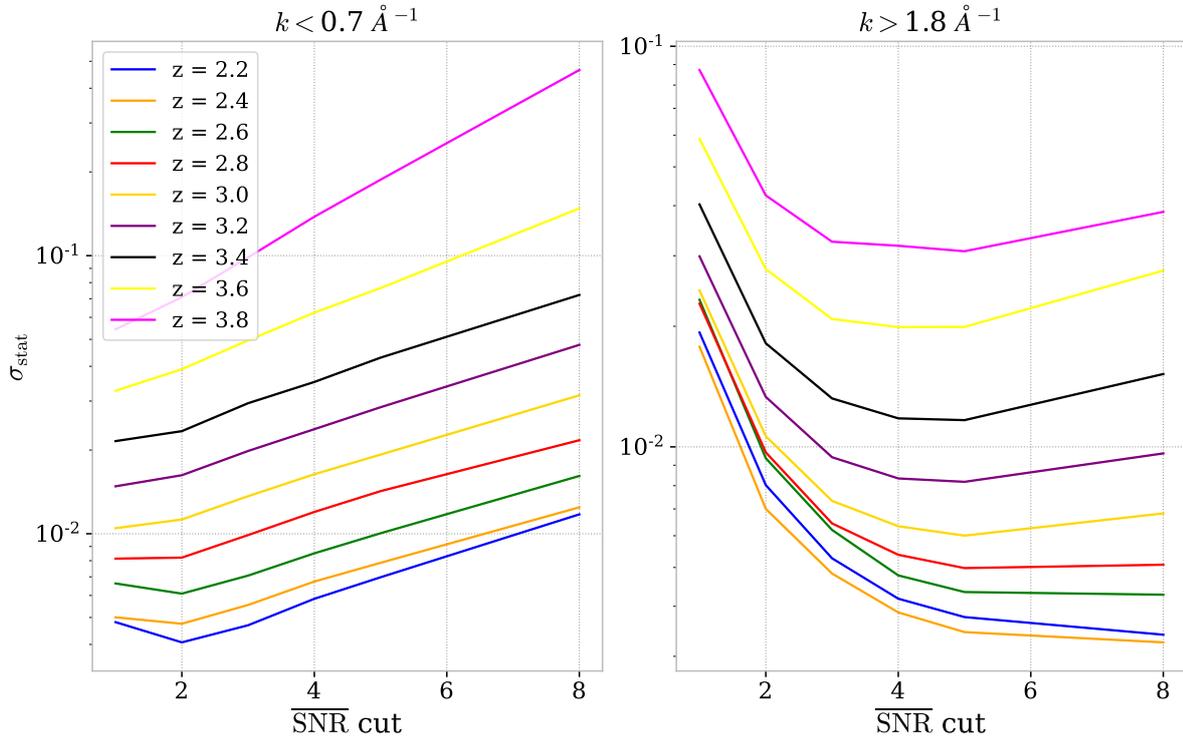


FIGURE 5.15 – Statistical uncertainties of the  $P_{1D,\alpha}$  measurement on SV3+DA0.2 in  $\text{\AA}$ , as a function of the  $\overline{\text{SNR}}$  cut. These uncertainties are calculated as the rms of measurements from individual  $P_{1D,\alpha}$  for each redshift bin. Here, we average statistical uncertainties over large scales ( $k < 0.7 \text{ \AA}^{-1}$ , left), and small scales ( $k > 1.8 \text{ \AA}^{-1}$ , right).

Redshift bin	2.2	2.4	2.6	2.8	3.0	3.2	3.4	> 3.6
$\overline{\text{SNR}}$ threshold	4.1	3.9	3.6	3.2	2.9	2.6	2.2	2.0

TABLE 5.2 –  $\overline{\text{SNR}}$  threshold applied to the different redshift bins of the DESI SV3+DA0.2  $P_{1D,\alpha}$  measurement.

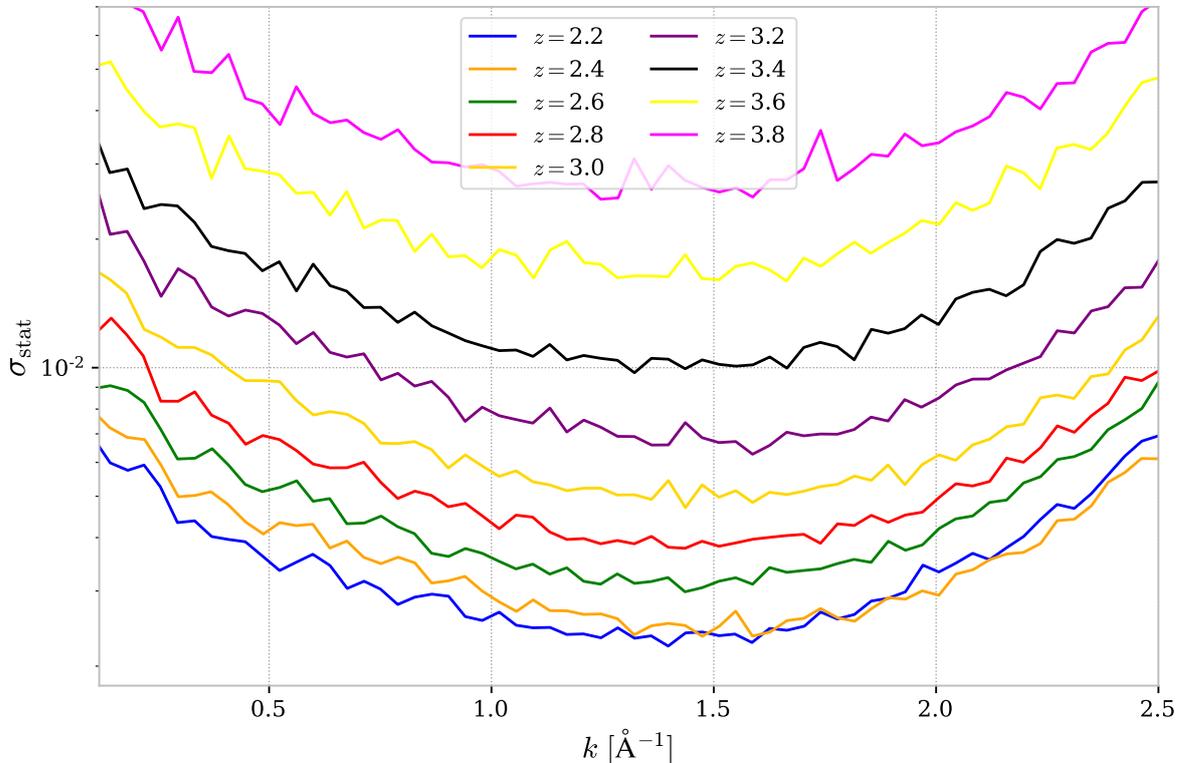


FIGURE 5.16 – Statistical uncertainties of the DESI SV3+DA0.2  $P_{1D,\alpha}$  measurement in  $\text{\AA}$ , as a function of wavenumber. These uncertainties are given with the redshift-dependent  $\overline{\text{SNR}}$  cut reported in Tab. 5.2.

$P_{1D,\alpha}$ . At large scales ( $k \lesssim 1 \text{ \AA}^{-1}$ ),  $\sigma_{\text{stat}}$  is a decreasing function of  $k$ , mainly due to the decrease in Fourier modes available to compute  $P_{1D,\alpha}$ .

### Systematic uncertainties

We associate systematic errors, noted  $\sigma_{\text{syst}}$ , to a number of effects studied in the previous sections for the  $P_{1D,\alpha}$  measurement. As this is a first measurement, we decide to make similar choices than [3] for the definitions of those systematic uncertainties:

- **Noise estimation:** As presented in Sec. 5.2.1, the pipeline noise is corrected using the  $P_{\text{noise,miss}}$  term. This correction depends on the data set considered. We assign a systematic uncertainty equal to 30 % of the average  $P_{\text{noise,miss}}$  for each redshift bin.
- **Side-band:** The fitted side-band power spectrum  $P_{\text{SB1,m}}$  measured in Sec. 5.2.2 is subtracted to  $P_{1D,\alpha}$  to account essentially for metal absorptions in the Ly $\alpha$  forest region. We associate to this correction a systematic uncertainty equal to 30 % of the measured SB1 power spectrum statistical errors.
- **Pixel masking:** The impact of masking HCD objects and sky lines on the  $P_{1D,\alpha}$  measurement was determined with mocks in Sec. 5.3.1. Pixel masking is corrected by multiplying

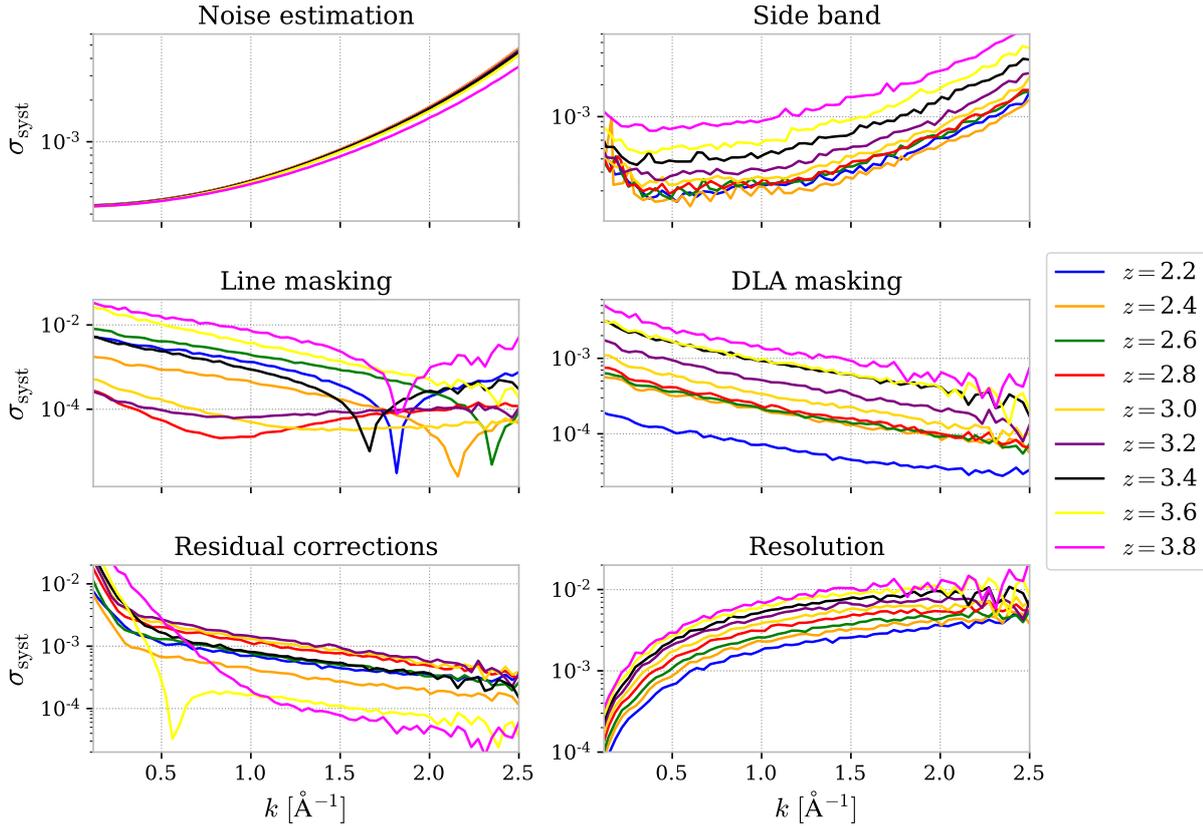


FIGURE 5.17 – Absolute value of the systematic uncertainties  $\sigma_{\text{syst}}$  in  $\text{\AA}$ , for different redshift bins, on the DESI SV3+DA0.2  $P_{1D,\alpha}$  measurement.

$A_{\text{sky}}(z, k) \cdot A_{\text{hcd}}(z, k)$  to the  $P_{1D,\alpha}$  estimator. We define the systematic error associated with each masking as 30 % of this correction.

- **Residual corrections:** We also assign a systematic error of 30 % times the  $A_{\text{residual}}(z, k)$  correction computed in Sec. 5.3.2.
- **Resolution:** We fit the average resolution correction given in Fig. 5.3 by a simplified model  $\exp(-0.5(k\Delta\lambda)^2) \cdot \text{sinc}(0.5k\Delta\lambda_{\text{pix}})$  with  $\Delta\lambda_{\text{pix}}$  fixed to the DESI pixel separation. This procedure allows to determine the effective resolution power  $\Delta\lambda$  in  $\text{\AA}$ . Measurements of DESI’s PSF stability, shown in [20] (Fig. 30), indicate that its fractional change is less than 1 % over all spectrographs. We therefore assign a conservative systematic uncertainty  $\sigma_{\Delta\lambda} = 1\% \Delta\lambda$ . Using the abovementioned simplified resolution model, this translates into a  $P_{1D,\alpha}$  uncertainty equal to  $2k^2\Delta\lambda\sigma_{\Delta\lambda} \cdot P_{1D,\alpha}(k)$ .

Fig. 5.17 shows the systematic uncertainties for different redshift bins, and Fig. 5.18 their relative values with respect to statistical errors. The general trends are similar to those of the SDSS measurement [3]. Note that, in our study, we did not yet consider the impact of HCD and BAL completeness which are beyond the scope of this thesis. Finally, we do not attempt to separate the different components of the residual corrections.

Given our limited statistics, most of the systematic errors are smaller than the statistical uncertainties for all redshift bins and all scales. However, this is expected to change for future DESI measurements, which will offer unprecedented statistics. Most of the analysis presented in this chapter should then be improved to reduce the systematic error budget.

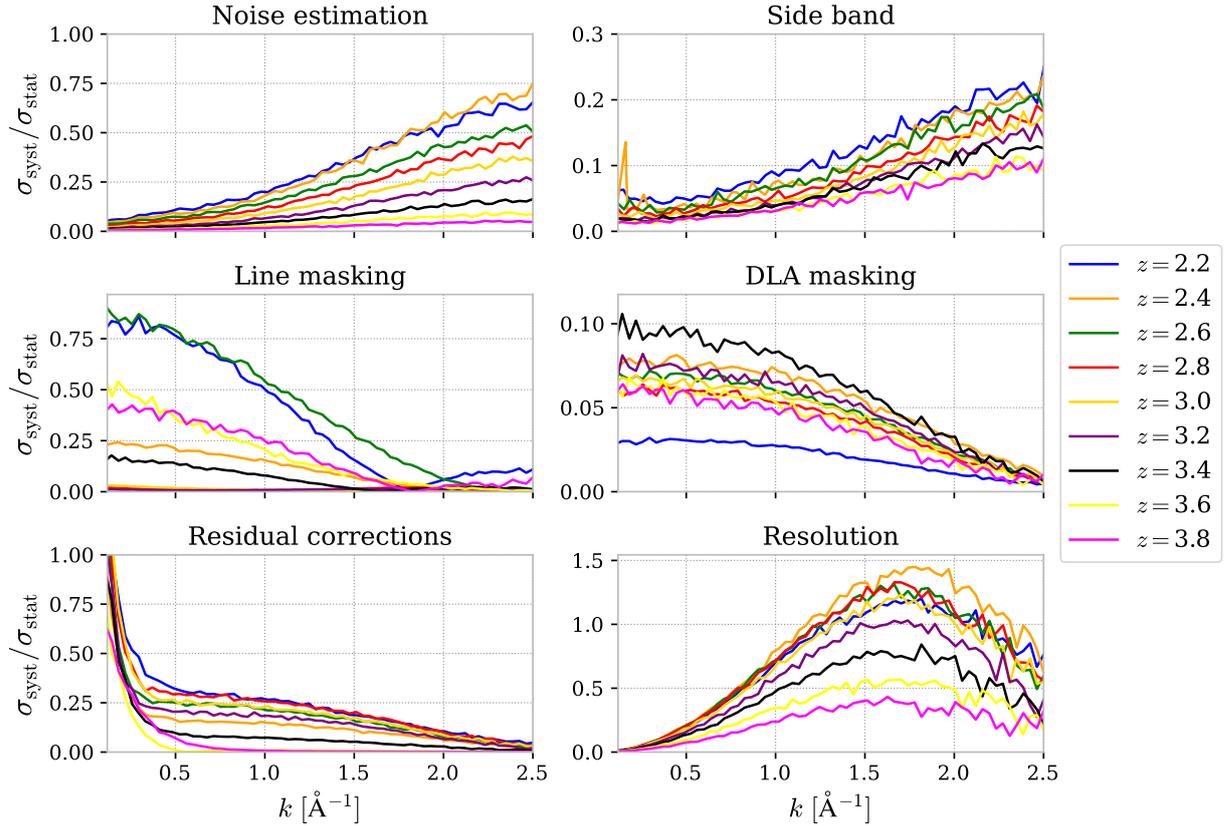


FIGURE 5.18 – Ratio between the systematic uncertainties shown in Fig. 5.17 and the statistical ones shown in Fig. 5.15, for different redshift bins.

## 5.4.2 Measurement

The  $P_{1D,\alpha}$  measurement is done with the pipeline presented in Sec. 5.1. More specifically, we reject QSOs whose balnicity index is larger than zero. We mask HCD objects detected by `desi-dlas`, with  $N_{\text{HI}} > 10^{20.3} \text{ cm}^{-2}$  (DLA), and confidence level higher than 70 %. We use the DESI sky emission line catalog ( $\Lambda_l = 2.5, \Lambda_w = 1.2$ )<sup>8</sup> for masking. Considering all the corrections defined in the previous section, the final  $P_{1D,\alpha}$  estimator is defined by

$$\begin{aligned}
 P_{1D,\alpha}(k) &= A_{\text{sky}}(z, k) \cdot A_{\text{hcd}}(z, k) \cdot A_{\text{residual}}(z, k) \cdot \\
 &\left( \left\langle [P_{\text{raw}}(k) - P_{\text{noise}}(k) - P_{\text{noise,miss}}(k)] \cdot \mathbf{R}^{-2}(k) \cdot \text{sinc}^2\left(\frac{k\Delta\lambda_{\text{pix}}}{2}\right) \right\rangle - P_{\text{SB1,m}}(k) \right).
 \end{aligned}
 \tag{5.34}$$

Fig. 5.19 presents this  $P_{1D,\alpha}$  measurement for 9 redshift bins ranging from 2.2 to 3.8, for wavenumbers  $0.118 \leq k \leq 2.5 \text{ \AA}^{-1}$ , and using a total of 21,050 sub-forests. The represented error bars are the statistical and systematic uncertainties added in quadrature.

This measurement constitutes the first  $P_{1D,\alpha}$  estimation at such spectral resolution with a large number of Ly $\alpha$  forests. We are able to reach the wavenumber  $k_{\text{max}} = 2.5 \text{ \AA}^{-1}$  for all redshift bins. In comparison, the eBOSS measurement [3] achieved maximal wavenumber of  $k_{\text{max}} = 1.54 \text{ \AA}^{-1}$  for  $z = 2.2$  and  $k_{\text{max}} = 1.03 \text{ \AA}^{-1}$  for  $z = 3.8$ . This measurement was possible thanks to the first characterization of the DESI instrument and data, essential for future measurements.

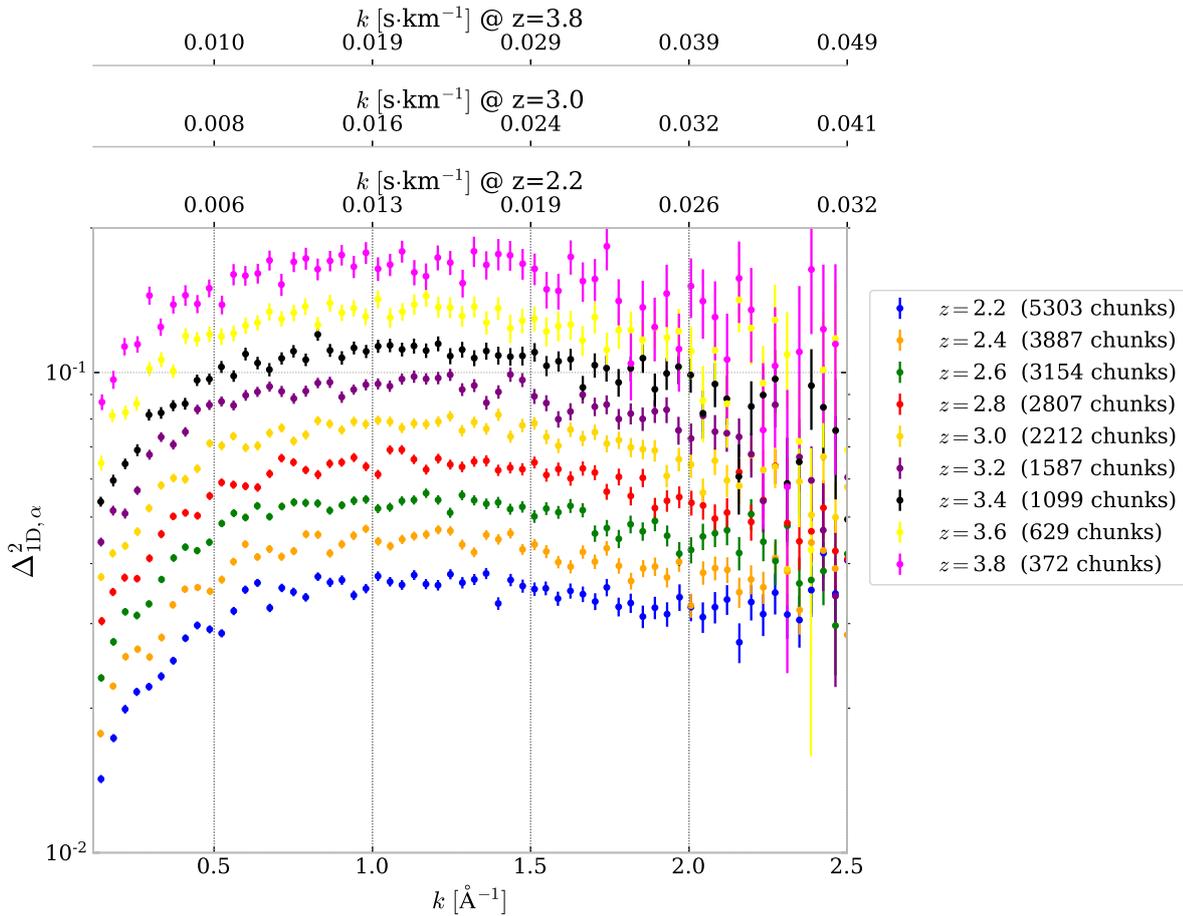


FIGURE 5.19 – Normalized one-dimensional Ly $\alpha$  forest power spectrum ( $\Delta_{\text{ID},\alpha}(k)$ ) using the SV3+DA0.2 data set, for redshift bins from  $z = 2.2$  to  $z = 3.8$ . All the corrections given in Eqn. 5.34, and the  $\overline{\text{SNR}}$  cuts reported in Tab. 5.2 are applied to realize this measurement. As an illustration, wavenumbers in velocity space for different redshifts are represented at the top of the figure. Error bars are the quadrature addition of systematic and statistical uncertainties.

The DESI SV3+DA0.2 data set remains inferior in statistics (94,558 sub-forests used for eBOSS [3]), but we expect that the end of the DESI survey will provide a considerable number of high-quality Ly $\alpha$  forests.

### 5.4.3 Summary and prospects

In this chapter, we have set up the building bricks necessary for a precision measurement of  $P_{\text{ID},\alpha}$  with DESI, and we presented a first preliminary measurement with SV3+DA0.2 data.

This work already demonstrates significant improvements with respect to eBOSS results. Tab. 5.3 summarizes the progress already achieved with respect to eBOSS, and items for which there is still room for improvement.

The next steps toward a first DESI  $P_{\text{ID},\alpha}$  measurement, not covered in this manuscript, will be to study the completeness of HCD and BAL removal and get a full understanding of the continuum fitting impact with mocks. We shall then be able to compare DESI measurement with published eBOSS [3] and high-resolution [5] measurements.

	eBOSS [3]	DESI (this work)	Comments
<b>Noise estimation</b>	Exposure difference	Pipeline model	DESI outperforms eBOSS, as empirical asymptotic correction is minor and can be easily estimated.
<b>Resolution</b>	Gaussian kernel	Resolution matrix	High wavenumber DESI data are beyond eBOSS reach and already yield physically sound measurements.
<b>Side-bands (metals)</b>	Polynomial fit	Semi-physical model	The DESI side-band power spectrum is clearly improved with respect to eBOSS.
<b>Sky line masking</b>	-	Updated list of masks	The impact of sky line masking is less important for DESI than for eBOSS (except at $z = 2.2$ ).
<b>HCD catalog</b>	CNN estimation	Similar	Impact of HCD completeness still to be studied (expected to be similar to eBOSS).
<b>BAL catalog</b>	Absorption finder	Similar	Impact of BAL completeness still to be studied.
<b>Continuum fitting</b>	From mocks	Similar	Work in progress to understand and minimize its impact on the measured $P_{1D,\alpha}$ .

TABLE 5.3 – Qualitative comparison between this work and the eBOSS measurement [3].

The statistical precision of this early DESI measurement together with its covered range of wavenumbers are expected to be already of interest for cosmology. In the near future, we plan to use the  $P_{1D,\alpha}$  emulation grid presented in Sec. 3.2.3 to obtain cosmological constraints.

This measurement was performed using the FFT estimator presented in Sec. 5.1.3. This estimator has the disadvantage of being biased with respect to masking (DLAs, sky lines). In parallel, a measurement on the same data using the quadratic maximum likelihood estimator is realized within the DESI collaboration. This estimator developed in [5, 21] is by construction not biased with respect to masking but presents a greater instability concerning the modeling of the noise found by [2].

The research work developed in this chapter is part of the one-dimensional study of matter distribution with DESI Ly $\alpha$  forest sample. The continuation of this thesis is devoted to developing relatively recent methods using the three-dimensional information of the Ly $\alpha$  forest. As these studies are chronologically earlier, they were mainly performed on eBOSS data.

## Bibliography

- [1] N. Palanque-Delabrouille, C. Yèche, N. Schöneberg, J. Lesgourgues, M. Walther, S. Chabanier et al., *Hints, neutrino bounds and WDM constraints from SDSS DR14 Lyman- $\alpha$  and Planck full-survey data*, *Journal of Cosmology and Astroparticle Physics* **2020** (2020) 038.
- [2] N. Palanque-Delabrouille, C. Yèche, A. Borde, J.-M. L. Goff, G. Rossi, M. Viel et al., *The one-dimensional Ly-alpha forest power spectrum from BOSS*, *Astronomy & Astrophysics* **559** (2013) A85.
- [3] S. Chabanier, N. Palanque-Delabrouille, C. Yèche, J.-M. L. Goff, E. Armengaud, J. Bautista et al., *The one-dimensional power spectrum from the SDSS DR14 Ly $\alpha$  forests*, *Journal of Cosmology and Astroparticle Physics* **2019** (2019) 017.
- [4] P. McDonald, U. Seljak, S. Burles, D. J. Schlegel, D. H. Weinberg, D. Shih et al., *The Lyman-alpha Forest Power Spectrum from the Sloan Digital Sky Survey*, *The Astrophysical Journal Supplement Series* **163** (2006) 80.
- [5] N. G. Karaçaylı, N. Padmanabhan, A. Font-Ribera, V. Iršič, M. Walther, D. Brooks et al., *Optimal 1D Ly  $\alpha$  forest power spectrum estimation – II. KODIAQ, SQUAD, and XQ-100*, *Monthly Notices of the Royal Astronomical Society* **509** (2022) 2842.
- [6] E. Chaussidon, C. Yèche, N. Palanque-Delabrouille, D. M. Alexander, J. Yang, S. Ahlen et al., *Target Selection and Validation of DESI Quasars*, Aug., 2022. 10.48550/arXiv.2208.08511.
- [7] N. Busca and C. Balland, *QuasarNET: Human-level spectral classification and redshifting with Deep Neural Networks*, *arXiv:1808.09955 [astro-ph]* (2018) .
- [8] J. Farr, A. Font-Ribera and A. Pontzen, *Optimal strategies for identifying quasars in DESI*, *arXiv:2007.10348 [astro-ph]* (2020) .
- [9] D. M. Alexander, T. M. Davis, E. Chaussidon, V. A. Fawcett, A. X. Gonzalez-Morales, T.-W. Lan et al., *The DESI Survey Validation: Results from Visual Inspection of the Quasar Survey Spectra*, Aug., 2022. 10.48550/arXiv.2208.08517.
- [10] D. Parks, J. X. Prochaska, S. Dong and Z. Cai, *Deep Learning of Quasar Spectra to Discover and Characterize Damped Lya Systems*, *arXiv:1709.04962 [astro-ph]* (2017) .
- [11] J. E. Bautista, N. G. Busca, J. Guy, J. Rich, M. Blomqvist, H. d. M. d. Bourboux et al., *Measurement of BAO correlations at  $z=2.3$  with SDSS DR12 Ly $\alpha$ -Forests*, *Astronomy & Astrophysics* **603** (2017) A12.
- [12] S. Chabanier, T. Etourneau, J.-M. L. Goff, J. Rich, J. Stermer, B. Abolfathi et al., *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: The Damped Lyman- $\alpha$  systems Catalog*, *arXiv:2107.09612 [astro-ph, physics:physics]* (2021) .
- [13] H. d. M. d. Bourboux, J. Rich, A. Font-Ribera, V. d. S. Agathe, J. Farr, T. Etourneau et al., *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Baryon acoustic oscillations with Lyman- $\alpha$  forests*, *The Astrophysical Journal* **901** (2020) 153.

- [14] H. du Mas des Bourboux, J. Rich, A. Font-Ribera, V. de Sainte Agathe, J. Farr, T. Etourneau et al., *picca: Package for Igm Cosmological-Correlations Analyses*, *Astrophysics Source Code Library* (2021) ascl:2106.018.
- [15] R. A. C. Croft, D. H. Weinberg, N. Katz and L. Hernquist, *Recovery of the Power Spectrum of Mass Fluctuations from Observations of the Lyman-alpha Forest*, *The Astrophysical Journal* **495** (1998) 44.
- [16] S. Bajtlik, R. C. Duncan and J. P. Ostriker, *Quasar Ionization of Lyman-Alpha Clouds: The Proximity Effect, a Probe of the Ultraviolet Background at High Redshift*, *The Astrophysical Journal* **327** (1988) 570.
- [17] M. Walther, J. F. Hennawi, H. Hiss, J. Oñorbe, K.-G. Lee, A. Rorai et al., *A New Precision Measurement of the Small-scale Line-of-sight Power Spectrum of the Ly $\alpha$  Forest*, *The Astrophysical Journal* **852** (2018) 22.
- [18] A. Kramida, Yu. Ralchenko, J. Reader and and NIST ASD Team. NIST Atomic Spectra Database (ver. 5.9), [Online]. Available: <https://physics.nist.gov/asd> [2022, May 17]. National Institute of Standards and Technology, Gaithersburg, MD., 2021.
- [19] K.-G. Lee, S. Bailey, L. E. Bartsch, W. Carithers, K. S. Dawson, D. Kirkby et al., *The BOSS Lyman-alpha Forest Sample from SDSS Data Release 9*, *The Astronomical Journal* **145** (2013) 69.
- [20] B. Abareshi, J. Aguilar, S. Ahlen, S. Alam, D. M. Alexander, R. Alfarsy et al., *Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument*, Tech. Rep. arXiv:2205.10939, arXiv, May, 2022. 10.48550/arXiv.2205.10939.
- [21] N. G. Karaçaylı, A. Font-Ribera and N. Padmanabhan, *Optimal 1D Ly- $\alpha$  Forest Power Spectrum Estimation I: DESI-Lite Spectra*, *Monthly Notices of the Royal Astronomical Society* **497** (2020) 4742.



# 6

---

## Lyman- $\alpha$ tomography

---

*“Une nébuleuse n’est pas un unique et énorme soleil, mais un système de nombreux soleils, rassemblés en raison de leur distance dans un espace si étroit, que leur lumière, qui serait imperceptible pour chacun d’eux isolément, parvient, grâce à leur innombrable quantité, à produire une blancheur pâle et uniforme.”*

– Emmanuel Kant, *Histoire générale de la nature et théorie du ciel*, 1755

---

**Contents**

---

<b>6.1</b>	<b>eBOSS data and tomographic method</b>	<b>201</b>
6.1.1	Stripe 82 eBOSS data	201
6.1.2	Wiener filter	204
<b>6.2</b>	<b>Map-making results</b>	<b>208</b>
6.2.1	Mocks	208
6.2.2	Matter density field of mocks	209
6.2.3	eBOSS data	211
<b>6.3</b>	<b>Applications</b>	<b>215</b>
6.3.1	Void finder	215
6.3.2	Catalog stacking	217
6.3.2.1	Void stack	217
6.3.2.2	QSO stacking	220
6.3.3	Matter overdensities	221
<b>6.4</b>	<b>DESI Lyman-break galaxies secondary target program</b>	<b>223</b>
	<b>Bibliography</b>	<b>226</b>

---

WITH the one-dimensional power spectrum presented in previous chapter, the correlations of the Ly $\alpha$  forest are measured only along the lines-of-sight (also noted los). As explained in Sec. 2.3, several measurements, for example auto-correlation and cross-correlation, use the Ly $\alpha$  forest in three dimensions. During my thesis, I focused on constructing a 3-dimensional map from Ly $\alpha$  absorption data. This is feasible on scales comparable to the mean separation between lines-of-sight [1]. With this method, called Ly $\alpha$  tomography, one obtains a map tracing the distribution of matter from the information of the Ly $\alpha$  forest.

Such an endeavour was already performed by the COSMOS Lyman-Alpha Mapping And Tomography Observations (CLAMATO) survey [2, 3, 4, 5]. This program uses a dense set of spectroscopic observations from Lyman-break galaxies (LBGs) to reconstruct the cosmic web in a small area ( $0.2 \text{ deg}^2$ ). With a redshift range  $z = 2.05 - 2.55$ , this survey covered a comoving volume of  $34 \times 28 \times 438 h^{-3} \cdot \text{Mpc}^3$ . With the combined use of LBGs and QSOs, the achieved total line-of-sight density is  $1455 \text{ deg}^{-2}$ . The ultimate purpose of such high sampling programs is to map the cosmic web down to the scale of filaments, i.e., of the order of the Mpc. Accessing filament scales requires very dense measurements of those faint LBGs. This will be possible with future observations using e.g., the Extremely Large Telescope (ELT) [6]. The performance of the algorithm used by CLAMATO was assessed through cosmological simulations in [7, 8]. These studies determined, in particular, the map-making performance for different mean separations ( $d_{\perp}$ ) between individual lines-of-sight.

In this chapter, I present a high-redshift Ly $\alpha$  forest absorption map on much larger scales, from part of the quasar spectra obtained by the Sloan Digital Sky Survey-IV (SDSS-IV) [9]. They are provided by the 16<sup>th</sup> Data Release (DR16) [10] of the SDSS extended Baryon Oscillation Spectroscopic Survey (eBOSS). I use the same tomographic reconstruction algorithm developed by CLAMATO. The authors in [7] also exploited BOSS DR9 data to obtain a tomographic map, but with a lower line-of-sight density than in this study. In my thesis, I focus on the Stripe 82 part of the SDSS footprint (right ascension from  $317^{\circ}$  to  $45^{\circ}$  and declination from  $-1.25^{\circ}$  to  $+1.25^{\circ}$ ). These data, as well as the method of Ly $\alpha$  tomography used in this study, will be presented in Sec. 6.1. The resulting tomographic map and its quality assessment using dedicated mocks are presented in Sec. 6.2.

Ly $\alpha$  forest tomographic mapping is already used for several applications, such as identifying underdensities in the matter distribution of the Universe. Large voids can be detected in the cosmic web by applying a void finder to the tomographic map. While several methods were developed to detect voids from galaxy surveys (e.g. VIDE algorithm [11]), detection based on the Ly $\alpha$  forest is less explored. It is possible to search for voids from the one-dimensional Ly $\alpha$  forest data, as shown in [12], but void searches based on 3D maps are more efficient and robust, making use of redundant information from nearby lines-of-sight. A void search method based on tomographic Ly $\alpha$  forest mapping was developed in [13] and applied in [14] to produce a void catalog from the cosmic web. In this chapter, I apply a similar method to produce a catalog of voids from the Stripe 82 tomographic map. Given the volume and resolution of this map, I can retrieve large voids in the cosmic web. This study is detailed in Sec. 6.3.

The Ly $\alpha$  tomography can also be used to study structures associated to overdensities in the cosmic web. It was applied to detect an enormous Ly $\alpha$  nebula [15] centered on the Ly $\alpha$  forest protoclusters found by the MAMMOTH project [16, 17]. Potential protoclusters are of course a target of choice for Ly $\alpha$  tomography. As an example, a  $z = 2.45$  protocluster was discovered in [18] and later confirmed as an overdensity of spectroscopically-observed galaxies. I perform a similar search with the tomographic map obtained in this thesis and provide the most probable

candidates for extensive protocluster systems in the Stripe 82 field. The results of this study are given in Sec. 6.3.

I also applied this tomographic method to the first DESI data used for the P1D measurement in Chap. 5. In parallel, I co-led a DESI secondary observation project which was accepted: its goal is to observe the distant Lyman break galaxies (LBGs) that could be targeted more intensively during future stages of DESI. My interest is to observe the LBG Ly $\alpha$  forest to increase the target density for Ly $\alpha$  tomography. For this project, we obtained a total exposure of 8 hours over the COSMOS field. The results of the efforts concerning Ly $\alpha$  tomography on DESI are reported in Sec. 6.4.

This chapter is largely based on my published article given in [19]. Changes with respect to the article mainly concern small improvement of the mocks, which are used for Chap. 7. For this analysis, I developed the python package `lelantos`<sup>1</sup>. All algorithms used in this chapter are documented in this public package.

---

1. LELANTOS - Large-scale LyMAN-alpha TOMography Suite <https://github.com/corentinravoux/lelantos/>

## 6.1 eBOSS data and tomographic method

In this section, I present the data used in my eBOSS tomographic reconstruction [19]. By adapting the algorithm used by CLAMATO to these data, it is possible to build a large-scale Ly $\alpha$  tomographic map.

### 6.1.1 Stripe 82 eBOSS data

To interpolate the lines-of-sight, it is preferable to work on a footprint over which observations are as uniform and dense as possible. The density drives the measurement of transverse variations of the absorption field. Homogeneity facilitates the creation, and subsequent analysis, of the map by using algorithms with the same parameters for the whole map.

#### Stripe 82 properties

In [19], I used the Ly $\alpha$  forest region from quasar spectra available in the 16<sup>th</sup> Data Release [10] (DR16) of the SDSS-IV eBOSS survey [9]. This survey as well as the telescope used were presented in Sec. 4.2.

The sky density of these objects, which is much smaller than that of LBGs, is the main factor limiting the resolution of my tomographic map. Compared to CLAMATO, which used QSOs and LBGs, the eBOSS QSO catalog provides a lower density.

For this reason, while the entire eBOSS footprint could be used in principle, I choose to focus on the **Stripe 82** field, a narrow band in the Equatorial plane lying within the Southern Galactic Cap. A particularly dense and uniform sample of quasar spectra was acquired in a large fraction of this field. Indeed, repeated photometric observations over about ten years lead to targeting quasars and rejecting stellar contaminants in a particularly efficient way. This efficient selection was accomplished by combining the time-variability information from observed lightcurves together with color measurements [20, 21]. This improved targeting efficiency brings the total quasar density with  $z \geq 2.1$  in the Stripe 82 field to 37 deg<sup>-2</sup>. In comparison, in the same redshift range, the total quasar density of the whole eBOSS footprint for DR16 is approximately 20 deg<sup>-2</sup>.

The footprint of my map within Stripe 82 is illustrated in Fig. 6.1. The right ascension range with a homogeneous Ly $\alpha$  forest dataset extends from -43° to +45°, and the declination range covers the whole Stripe 82 width, from -1.25° to +1.25°. This solid angle corresponds to 220 deg<sup>2</sup>. A total of 8199 Ly $\alpha$  forest spectra were obtained from quasar observations in this field, corresponding to the total density of 37 deg<sup>-2</sup>. The **total density** is defined here as the mean quasar density over a footprint (without considering the redshift direction). Lines-of-sight are selected over the redshift range  $z = 2.1 - 3.2$ .

The **effective density** of line-of-sight data depends in practice on the redshift because the lines-of-sight have a fixed length which covers redshifts between Ly $\alpha$  and Ly $\beta$  emission lines (see Chap. 2). Using an algorithm I developed in `lelantos`<sup>1</sup>, I made bins in redshift and for each bin calculated the line-of-sight density contributing to the corresponding redshift. I also calculated the **mean separation between nearest lines-of-sight**  $\langle d_{\perp} \rangle$  defined by

$$\langle d_{\perp} \rangle(z) = \frac{1}{N_{\text{los}}} \sum_{i=1}^{N_{\text{los}}} \min_{j \in [1, N_{\text{los}}], j \neq i} \left( d_{\perp}^{ij}(z) \right), \quad (6.1)$$

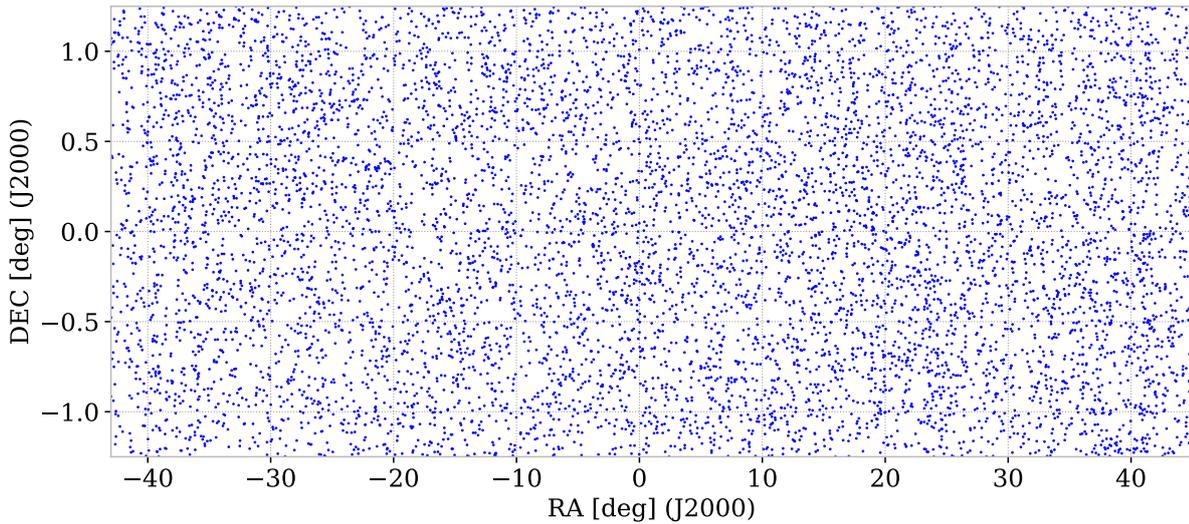


FIGURE 6.1 – Sky distribution of the 8199 eBOSS - DR16 quasars in the Stripe 82 ( $z = 2.1 - 3.2$ ), which provides Ly $\alpha$  forest spectra for this study. The total density is  $37 \text{ deg}^{-2}$ . The scales of the horizontal and vertical axes are different.

where  $N_{\text{los}}$  is the number of lines-of-sight and  $d_{\perp}^{ij}(z)$  is the comoving transverse distance between lines-of-sight  $i$  and  $j$  at a redshift  $z$ .

The effective density and the mean separation between nearest lines-of-sight as a function of redshift for Stripe 82 are shown in Fig. 6.2.

### Ly $\alpha$ contrast extraction

The quasar spectra used originate from the SDSS DR16 Quasar catalog [22]. Quasar targets were observed over the five-year period during the BOSS survey [23, 24] or during the SDSS-IV eBOSS survey [9, 25].

As explained in Sec. 4.2.3, the SDSS pipeline is used to coadd individual observations from typically four 15-minute long exposures. In the end, an unbiased estimator for the observed flux  $f(\lambda)$  is obtained.

For each spectrum, the Ly $\alpha$  contrast  $\delta_F$  is computed as detailed in Sec. 5.1.2. The Ly $\alpha$  contrast standard deviation  $\sigma_{\delta_F}$  is defined by Eqn. 5.9 where  $\sigma_{\text{pip,q}}$  is provided by the SDSS pipeline as detailed in Sec. 4.2.3. The corresponding signal-to-noise ratio per pixel (defined in Sec. 4.1.1) is typically close to 2. This value is relatively large compared to the case of CLAMATO. The main reason for this difference is that SDSS quasars are considerably brighter than LBGs.

The Ly $\alpha$  contrast extraction is realized with different parameters than for  $P_{1D,\alpha}$ . First, all the quasars available are used for the continuum fitting without applying a signal-to-noise ratio cut. For Ly $\alpha$  tomography, the Ly $\alpha$  contrast calculation is restricted to pixels for which the observed wavelength is in the range  $3769 - 5106 \text{ \AA}$ , corresponding to  $2.1 < z < 3.2$ . The same cut are applied for rest-frame wavelength, which implies a redshift difference  $\Delta z = 0.041$  between the quasar and the end of the line-of-sight. When distances are computed as detailed in Sec. 6.1.2, this separation from the quasar is  $49 h^{-1} \text{ Mpc}$ , which is sufficient to suppress the proximity effect. All the terms for the continuum fitting in Eqn. 5.7 are used and the polynomial order of continuum in Eqn. 5.4 is one. The number of continuum fitting iterations is set to 5 after

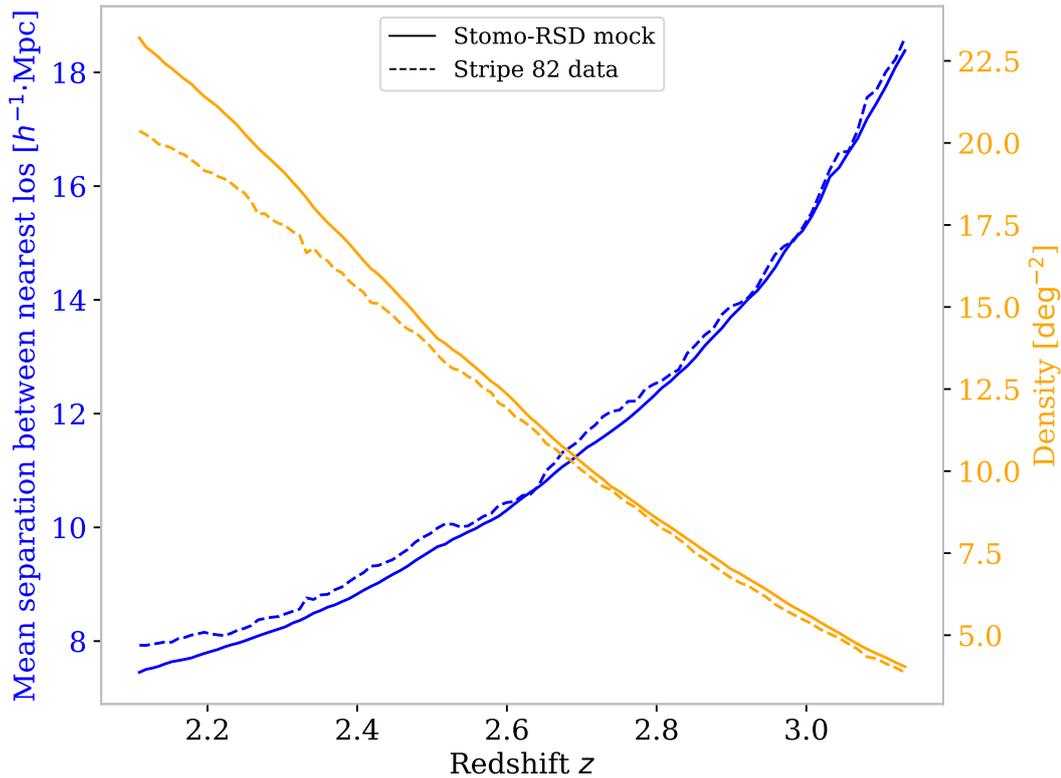


FIGURE 6.2 – (Blue) Mean separation between nearest lines-of-sight [ $h^{-1}\cdot\text{Mpc}$ ] as a function of redshift. (Orange) Effective line-of-sight density [ $\text{deg}^{-2}$ ], as a function of redshift. DR16 data are shown in continuous curves while dashed curves correspond to *Stomo-RSD* mock presented in Sec. 6.2.1. The higher difference at low redshift comes from high-noise quasars, which are less contributing to the tomographic reconstruction. In comparison to [19], the mocks are improved to better mimic eBOSS data.

verifying continuum convergence. Finally, the wavelength binning is logarithmic here ( $\Lambda = \log(\lambda)$  in Eqn. 5.5), and no rebinning is applied either in rest or observed frames.

### Contaminants

As for the calculation of the one-dimensional Ly $\alpha$  power spectrum in Chap. 5, the spectra are contaminated by Ly $\alpha$  forest systematics detailed in Sec. 2.1.3.

Similarly to  $P_{1D,\alpha}$  calculation, I remove BAL QSO whose balnicity index is larger than zero. This suppresses 6% of quasars and is not counted in the 8199 quasars. HCD objects contaminate the Ly $\alpha$  contrast and I mask the affected regions of the spectra as it is done for the  $P_{1D,\alpha}$  measurement. I use the DR16 HCD catalog [26] obtained with the DLA finder `dla_cmn` [27]. This algorithm is a previous version of the `desi-dlas`<sup>4</sup> code, and was trained on visually inspected SDSS spectra. For Ly $\alpha$  tomography, I removed HCD candidates whose column density is  $N_{\text{HI}} > 10^{19.7} \text{ cm}^{-2}$  with a detection confidence larger than 70 %. This limit corresponds to most DLAs, and some sLLS (as defined in Sec. 2.1.3). In practice, for my data set the HCD detection efficiency is poor for  $N_{\text{HI}} < 10^{20.3} \text{ cm}^{-2}$  as mentioned in [26]. The fraction of Ly $\alpha$  forest pixels lost by this cut is 11 %. Finally, following [28], I mask the sky lines defined by the corresponding eBOSS catalog<sup>2</sup>.

2. [https://github.com/igmhub/picca/blob/master/etc/list\\_veto\\_line\\_Pk1D.txt](https://github.com/igmhub/picca/blob/master/etc/list_veto_line_Pk1D.txt)

The resulting set of  $(\delta_F, \sigma_{\delta_F})$ , as a function of sky coordinates (right ascension RA, declination DEC) and redshift  $z$ , constitutes the input data to the next step of tomographic map-making.

## 6.1.2 Wiener filter

Tomographic reconstruction algorithms can be used to infer the 3-dimensional IGM density from a set of 1-dimensional Ly $\alpha$  forest lines-of-sight. In [29, 30], several techniques are presented to adapt tomographic methods to the creation of a 3D Ly $\alpha$  contrast. In [29], a Bayesian method is presented. Advanced use of Bayesian methods in this context is also given in [31] and [32, 33], which apply a dynamic forward modeling approach to model the observed 1D lines-of-sight and reconstruct the 3D density and velocity fields of the IGM.

Under certain hypotheses detailed in [29], the Bayesian approach is mathematically equivalent to applying a **Wiener filter** [34, 35]. Wiener filtering provides an unbiased minimum-variance linear estimator of a field. A simple implementation of this method, adapted to Ly $\alpha$  forest datasets, was developed by CLAMATO [2, 3, 4, 5, 13, 36]; and my work uses the same public code, DACHSHUND<sup>3</sup>. An advantage of this approach is its technical simplicity: it does not rely on complex assumptions for the underlying model, but still includes the effects of instrumental properties of the input data.

Since the average separation between nearest lines-of-sight for SDSS data is around  $10 h^{-1}$ .Mpc, I reconstruct a map with a resolution close to this value. Therefore, only large-scale Ly $\alpha$  or matter contrasts are preserved in the reconstructed map. Their distribution is nearly Gaussian, so the underlying linear model used for Wiener filtering is well adapted to my case. The authors in [4, 5] found that Wiener filtering, with adapted parameters, is appropriate even in the case of the CLAMATO survey whose line-of-sight density is much larger than that of eBOSS.

### Wiener filter framework

The Wiener filter takes as an input the 1D line-of-sight data  $\vec{d}$ , assumed to be the sum of pixel signals  $\vec{s}_p$  and a supposed Gaussian noise  $\vec{n}$ :

$$\vec{d} = \vec{s}_p + \vec{n}. \quad (6.2)$$

It is interesting to note that this inverse problem is formally close to the spectral extraction of DESI detailed in Sec. 4.3.4. In the case of the Ly $\alpha$  tomography, it is not an estimate of  $\vec{s}_p$  that we aim to obtain but the reconstructed perfect signal on a map  $\vec{s}_m$  interpolated from  $\vec{s}_p$ .

The goal of the Wiener filter is to obtain an estimator  $\hat{\vec{s}}$  of  $\vec{s}_m$  for the Ly $\alpha$  contrast signal over the entire volume. For this purpose, the Wiener filter minimizes the difference between these two quantities

$$\epsilon = E \left[ \left| \vec{s}_m - \hat{\vec{s}} \right|^2 \right]. \quad (6.3)$$

To perform this minimization, I assume that the covariance between the map signal  $\vec{s}_m$  and the Gaussian noise is zero such that  $E[\vec{s}_m \vec{n}^t] = 0$ . The details of this calculation are given in [29, 36] and provide the following minimal error estimator:

$$\hat{\vec{s}} = \left( \mathbf{S}_{mp} (\mathbf{S}_{pp} + \mathbf{N})^{-1} \right) \cdot \vec{d}, \quad (6.4)$$

---

3. <https://github.com/caseywstark/dachshund>

where  $\mathbf{S}_{\text{mp}} = E [\vec{s}_{\text{m}} \vec{s}_{\text{p}}^{\text{t}}]$  is the map-pixel covariance matrix,  $\mathbf{S}_{\text{pp}} = E [\vec{s}_{\text{p}} \vec{s}_{\text{p}}^{\text{t}}]$  is the pixel-pixel covariance matrix, and  $\mathbf{N} = E [\vec{n} \vec{n}^{\text{t}}]$  is the noise covariance matrix. The matrix  $\mathbf{S}_{\text{mp}}$  is inaccessible and its form need to be assumed to solve this problem. For the Wiener filter, the pixel-pixel and map-pixel covariance matrices  $\mathbf{S}_{\text{pp}}$  and  $\mathbf{S}_{\text{mp}}$  are both modeled by Gaussian kernels:

$$\mathbf{S}_{ij} = \sigma_F^2 \exp \left( -\frac{(r_{i\parallel} - r_{j\parallel})^2}{2L_{\parallel}^2} \right) \exp \left( -\frac{(r_{i\perp} - r_{j\perp})^2}{2L_{\perp}^2} \right), \quad (6.5)$$

where  $i$  and  $j$  indexes are replaced by p or m.  $L_{\perp}$  and  $L_{\parallel}$  are the **transverse and longitudinal correlation lengths** of the signal, which are discussed below. The parameter  $\sigma_F^2$  is the **expected variance** of the Ly $\alpha$  contrast in the map. For my data, the noise matrix  $\mathbf{N}$  is well approximated by a diagonal matrix with  $\mathbf{N}_{ii} = n_i$ . The  $n_i$  are the values of  $\vec{n}$  and can be approximated by the estimate of the instrumental noise  $\sigma_{\delta_F}$ .

An estimate of the covariance of the map can be computed with the following estimator (detailed in [29, 36]):

$$\mathbf{M} = \mathbf{S}_{\text{mp}} \left( (\mathbf{S}_{\text{pp}} + \mathbf{N})^{-1} \mathbf{S}_{\text{pm}} \right), \quad (6.6)$$

where  $\mathbf{S}_{\text{pm}} = E [\vec{s}_{\text{p}} \vec{s}_{\text{m}}^{\text{t}}]$  is the pixel-map covariance matrix.

The DACHSHUND public code is an implementation of this method for reconstructing the Ly $\alpha$  forest over a map [36]. It uses a pre-conditioned conjugate gradient (PCG) [37] to inverse the matrices of the Wiener filter (Eqn. 6.5). PCG is an efficient numerical inversion method that quickly converges for sparse matrices.

### Coordinate transformation

The DACHSHUND software uses Cartesian input coordinates. Therefore, it is necessary to convert the positions of input data ( $\delta_F, \sigma_{\delta_F}$ ) into cartesian coordinates.

Using my `lelantos`<sup>1</sup> package, I converted the line-of-sight angular and redshift coordinates (RA, DEC,  $z$ ) into pixel coordinates ( $X, Y, Z$ ) in  $h^{-1}$ ·Mpc. I assume a  $\Lambda$ CDM model with a Planck cosmology [38], i.e., fixed  $\Omega_{\Lambda} = 0.6853$ ,  $\Omega_{\text{m}} = 0.3147$ ,  $\Omega_k = 0$ . Since all coordinates are expressed in  $h^{-1}$ ·Mpc, no assumption on the value of  $H_0$  is needed. As mentioned in [4, 5], the choice of cosmology only affects the global longitudinal and transverse scales.

To link the angular coordinates to Cartesian coordinates, we perform a spherical transformation after calculating the radial  $D_C$  and transverse  $D_M$  comoving distances given by the equations 1.29 and 1.30. As  $\Omega_k = 0$ , these two distances are equal ( $D_M = D_C$ ). The complete transformation that we will call **full** is given by

$$\begin{aligned} X &= D_C(z) \times \sin(\text{RA}) \times \cos(\text{DEC}), \\ Y &= D_C(z) \times \sin(\text{DEC}), \\ Z &= D_C(z) \times \cos(\text{RA}) \times \cos(\text{DEC}). \end{aligned} \quad (6.7)$$

On the other hand, it is technically convenient to define an alternative coordinate system such that all lines-of-sight are parallel to each other, and the output map is a rectangular box. In that case, the transverse comoving distances are computed for all pixels using the middle redshift ( $\bar{z} = 2.65$  in our case). This simpler transformation called **middle** is given by

$$\begin{aligned}
X &= D_C(\bar{z}) \times \text{RA}, \\
Y &= D_C(\bar{z}) \times \text{DEC}, \\
Z &= D_C(z).
\end{aligned} \tag{6.8}$$

This transformation is used for several applications developed in this chapter and Chap. 7. The spread in declination of the eBOSS Stripe 82 data is  $2.5^\circ$ . This implies that distance errors for this transformation between redshifts  $z = 2.1$  to  $3.2$  are few percent along the  $y$ -axis of the map. Along the right ascension direction,  $88^\circ$  wide, the difference is much larger as the parallel approximation does not hold. Therefore, a map build with the middle transformation is not adapted for, e.g., correlation studies on large distances ( $\gtrsim 50 h^{-1}\cdot\text{Mpc}$ ).

The DACHSHUND input is a list of pixel vectors  $(X, Y, Z, \delta_F, \sigma_{\delta_F})$ . They are used to construct  $r_{\parallel} = Z$ ,  $r_{\perp} = \sqrt{X^2 + Y^2}$ ,  $\vec{d}$ , and  $\mathbf{N}$  following the definition of the Wiener filter. The output is a 3D map containing the **reconstructed Ly $\alpha$  contrast**  $\delta_{F\text{map}}$  that corresponds to the signal  $\hat{\delta}$ .

### Parameter tuning

Wiener filter tomographic reconstruction is not a parameter-free method. The DACHSHUND software takes as input the following parameters:

- **Correlation lengths:** The transverse and longitudinal correlation lengths  $L_{\perp}$  and  $L_{\parallel}$  control the typical smoothing lengths of the reconstructed map.
- **Expected variance:** This parameter  $\sigma_F^2$  is related to the reconstructed Ly $\alpha$  contrast at the smoothing scales of the map. It controls the amplitude of fluctuations in the reconstructed map.
- **Convergence parameters:** Some parameters are used to optimize the convergence of the conjugate gradient algorithm. In this study they are taken the same as CLAMATO [4, 5]

The transverse correlation length  $L_{\perp}$  is constrained by the fact that no transverse Ly $\alpha$  contrast can be reconstructed with a size smaller than the mean separation between nearest lines-of-sight  $\langle d_{\perp} \rangle(z)$ . Fig. 6.2 shows that this quantity increases with redshift. It is related to the redshift distribution of quasar sources used in the sample. From this figure, I choose  $L_{\perp} = 13 h^{-1}\cdot\text{Mpc}$ , so that  $L_{\perp} > \langle d_{\perp} \rangle(z)$  for  $z < 2.8$ , which covers most of the map. At high redshift,  $z > 2.8$ ,  $L_{\perp}$  is smaller than the mean line-of-sight separation, a fact which must be considered when interpreting the map.

In principle, the longitudinal smoothing length  $L_{\parallel}$  could be set to a value close to  $1 h^{-1}\cdot\text{Mpc}$ , which corresponds to the spectrograph resolution. However, since this value is much smaller than  $\langle d_{\perp} \rangle$ , I choose to set  $L_{\parallel} = L_{\perp} = 13 h^{-1}\cdot\text{Mpc}$ . This choice reduces the amplitude of anisotropies in the Wiener-filtered map. It keeps large-scale information while removing small-scale longitudinal fluctuations of the Ly $\alpha$  contrast.

The maximal reconstructed density contrasts are controlled by the ratio  $\sigma_F^2/n_{\min}^2$ , where  $n_{\min}$  is the minimal value of the diagonal elements of the noise matrix  $\mathbf{N}$  [36]. Indeed, to avoid numerical divergences in the Wiener filter algorithm, these matrix elements must be bounded by a floor value. Given the statistical distribution of pixel noise in my dataset, I adopt  $n_{\min} = 0.1$ , so that for the 6.3 % of the pixels with the lowest noise, this minimum value is used rather than the true pixel noise for the corresponding elements  $n_i$ . In the CLAMATO analysis,  $n_{\min}$  was set to 0.2 due to larger relative noise in LBG spectra [4, 5]. As for CLAMATO, I choose  $\sigma_F^2$  such that the maximal contrast of the Wiener-filtered map is 1, which implies  $\sigma_F^2 = 0.01$ .

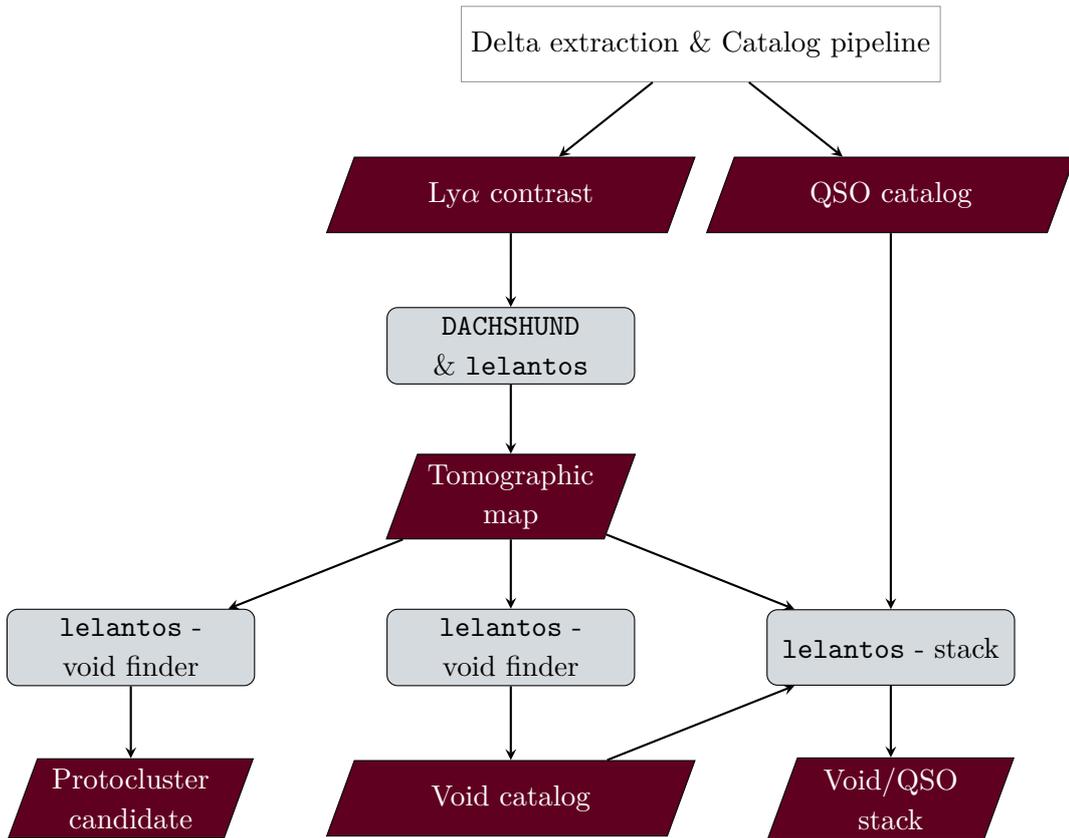


FIGURE 6.3 – Diagram of the  $\text{Ly}\alpha$  tomography pipeline. The delta extraction and catalog pipeline processes are different from those of the P1D pipeline in Fig. 5.1. In the case of  $\text{Ly}\alpha$  tomography, the delta extraction in `picca` is adapted to SDSS data with different parameters detailed in Sec. 6.1.1. Unlike the QSO catalog used in the pipeline in Fig. 5.1, the catalog used here is from [22].

The map is chosen to be pixelated at one pixel per  $h^{-1}\cdot\text{Mpc}$  for this chapter. Several tests related to the map’s properties showed that a map with half the resolution had the same properties.

Finally, since the number of pixels of the reconstructed map is much larger than for CLAMATO, I implemented a parallelized map-making method in `lelantos`. With this method, the map’s footprint is divided into chunks. The software `DACHSHUND` is launched simultaneously for each chunk, and the reconstructed sub-maps are merged. I let the chunks overlap to avoid edge effects, with a free parameter for the overlap width: I took the overlap width equal to the correlation length  $L_{\perp}$ . However, this parameter can be larger to minimize edge effects in very line-of-sight dense regions. During merging, the overlap regions are removed.

The full pipeline used for map-making and applications detailed in Sec. 6.3 is illustrated in Fig. 6.3.

## 6.2 Map-making results

To attest the quality of the **tomographic map of Ly $\alpha$  contrast**, tests were first performed on synthetic data called mock.

### 6.2.1 Mocks

To assess the performance of tomographic reconstruction, I do not use physical simulations of the Ly $\alpha$  forest, as they require hydrodynamical effects and a resolution on the order of  $100 h^{-1}\text{-kpc}$ . No such simulation exists covering a near-Gpc<sup>3</sup> volume for now. Instead, I use the `SaclayMocks` described in Sec. 3.1.1.

#### Mock creation

For the Ly $\alpha$  tomography, I adapted the `SaclayMocks` to match the sky density and redshift distribution of eBOSS-DR16 quasars in Stripe 82 described in Sec. 6.1.1. BAO studies like [39], for which these mocks were initially created, use the full length Ly $\alpha$  forest. Consequently, the lines-of-sight that finishes near  $z = 3.2$  were not used and not modeled. In the case of tomography, for high and low redshift, I used "partial" Ly $\alpha$  forests, i.e. lines-of-sight data whose redshift range extend beyond  $z = 3.2$  or below  $z = 2.1$ . Therefore, I had to increase the maximum redshift of the mocks.

The systematics are modeled using `quickquasars` [40] as shown in Sec. 3.1.1. For the specific case of Stripe 82, I created two mocks *Stomo-RSD* and *Stomo-noRSD* which both model the SDSS telescope noise, the QSO continuum, the continuum fitting of `picca` [41], the metal and HCD contamination, and the HCD detection and masking. As for the data detailed in Sec. 6.1.1, the continuum fitting and HCD detection are performed with the same algorithms. The only contamination not considered is that of the BAL QSOs. Therefore, I assume here that the detection of BALs in the data is perfect.

The difference between the two mocks *Stomo-RSD* and *Stomo-noRSD* is the presence or absence of RSD in the Ly $\alpha$  forest. RSD is removed in the mocks by forcing the  $c_{\text{GP}}$  term in Eqn. 3.9 to be zero for all redshifts. Except for differentiating the impact of RSD, all results presented from now on are from the *Stomo-RSD* mock.

The final outputs of the mocks are spectra and an associated quasar catalog, over a footprint matching my Stripe 82 selection. The mock spectra are then treated with the same pipeline and parameters as for the DR16 data detailed in Sec. 6.1.1.

#### Tomographic map

Fig. 6.2 illustrates the agreement between *SMocks-RSD* mock and data concerning the line-of-sight separation as a function of redshift. The histogram of noise,  $\sigma_{\delta_F}$ , is presented for *SMocks-RSD* mock and my selected data in Fig. 6.4.

The average and standard deviations of the noise distributions are similar, but their shapes are significantly different, especially at high noise. This result is related to the fact that there are fewer high magnitude quasars in the mocks than in the DR16 data. However, the difference in the high-noise tail has a minor impact on the tomographic map because Wiener filtering assigns more weight to low-noise pixels. I checked this effect explicitly, by constructing maps for which high-noise data were removed. This result is in agreement with [4, 36].

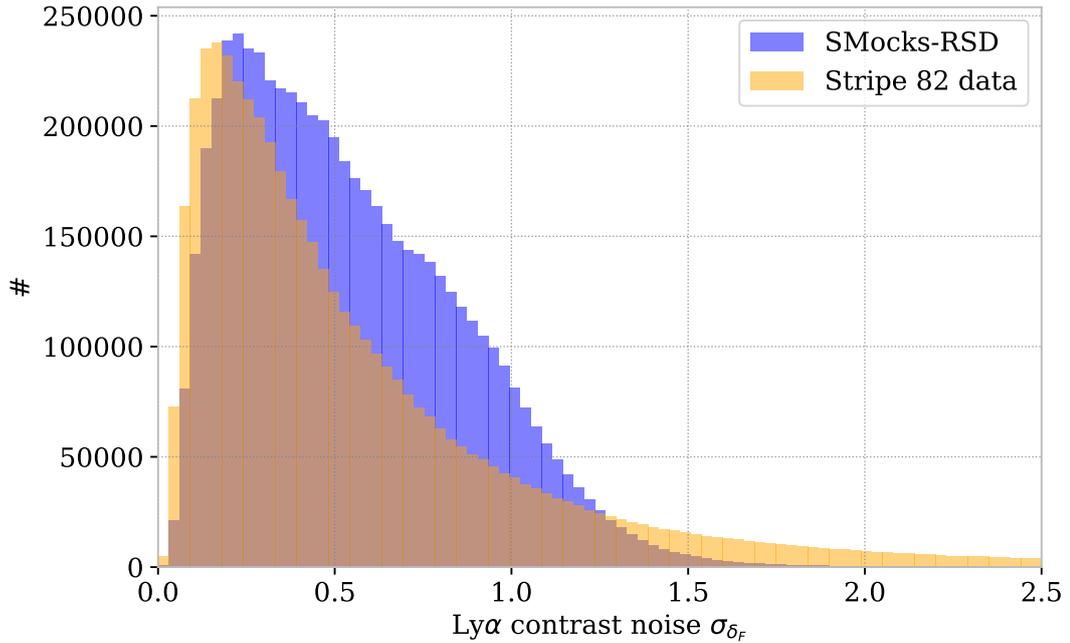


FIGURE 6.4 – Histograms of the Ly $\alpha$  contrast noise  $\sigma_{\delta_f}$ , for both our selected Stripe 82 data and *SMocks-RSD* mock. A pixel corresponds to the Ly $\alpha$  contrast over 1 Å width, implying equivalence between noise per pixel and per Å.

I apply the tomographic algorithm described in Sec. 6.1.2 to all mock data. Fig. 6.5 displays a slice of one quarter of the *SMocks-RSD* mock map. The reconstructed Ly $\alpha$  contrast,  $\delta_{F_{\text{map}}}$ , is represented by a color map such that matter overdensities, corresponding to negative Ly $\alpha$  contrast, appear in red, while matter underdensities are in blue. Map pixels located at a distance larger than  $20 h^{-1}\cdot\text{Mpc}$  from any line-of-sight are masked, as they can not be reasonably reconstructed; they appear in white in the representation. The process of Ly $\alpha$  tomography has also been performed for mocks with the full coordinate transformation defined in Sec. 6.1.2. As this map is not used for the applications shown in this thesis, it is not presented here.

## 6.2.2 Matter density field of mocks

Since the tomographic map measures Ly $\alpha$  contrast fluctuations on scales larger than  $13 h^{-1}\cdot\text{Mpc}$ , it is expected to trace the associated matter fluctuations on similar scales.

In order to estimate to what extent my tomographic map reproduces the underlying field of matter fluctuations, I measured its correlation with the underlying matter field in the mock. As shown in Sec. 3.1.1, it is possible to define a matter density field based on the initial Gaussian random field of the mocks. Eqn. 3.13 defines this matter field by taking into account the growth of structures. In practice, several steps are necessary to extract a matter field  $\delta_m$  of the *SacLayMocks* with the same coordinates as the associated tomographic map.

The large-scale matter field  $\delta_L$  is defined by Eqn. 3.1 and is expressed at redshift  $z = 0$  for *SacLayMocks*. The first step is to convert the angular and redshift coordinates related to the observation into coordinates in the matter box. I converted all the points of my tomographic map into the  $\delta_L$  box space using the middle conversion of Eqn. 6.8. As the pixel coordinates do not match immediately between the two boxes, a nearest-neighbor interpolation is done.

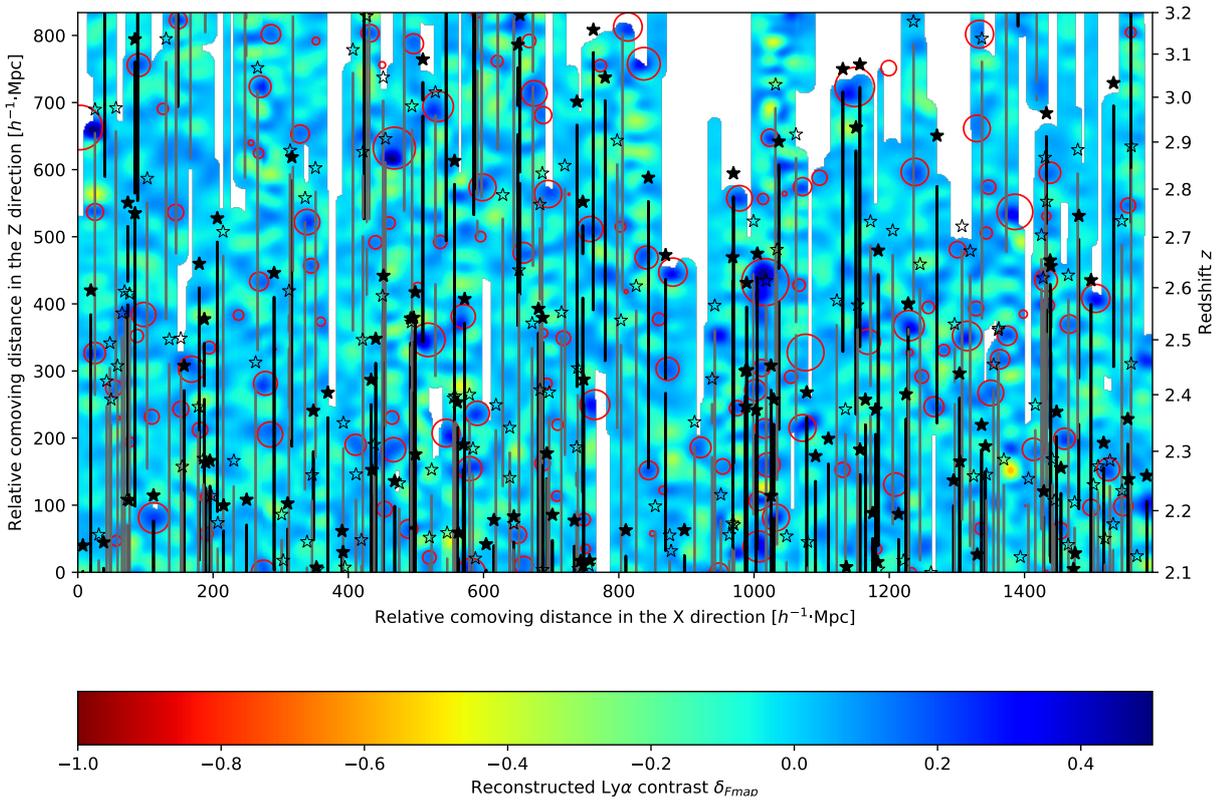


FIGURE 6.5 – Slice through the tomographic Ly $\alpha$  contrast map derived from the *SMocks-RSD* mock. It covers a  $22^\circ$  right-ascension range (along the  $x$  axis) at fixed declination. Circles are the intersection between the represented slice and identified voids from the *raw* catalog presented in Sec. 6.3.1. Filled (empty) stars represent QSOs whose distance along the  $y$  axis is less than 5 (10)  $h^{-1}\cdot\text{Mpc}$  away from the slice. Lines-of-sight used for the tomographic reconstruction, which are less than 5 (10)  $h^{-1}\cdot\text{Mpc}$  away from the slice, are pictured as black (grey) lines. Pixels located more than 20  $h^{-1}\cdot\text{Mpc}$  from any line-of-sight are masked.

Thanks to this transformation, I have recovered the large-scale matter field for each pixel of my tomographic map.

This map is transformed to a matter density field  $\delta_m$  using Eqn. 3.13 and can be Gaussian smoothed to represent the large-scale fluctuation of matter and to remove numerical effects. The matter density field in the box is then defined by

$$\delta_m = \delta_m^{\text{SMocks}} \otimes \text{Gauss}(d_{\text{filter}}), \quad (6.9)$$

where  $\delta_m^{\text{SMocks}}$  is defined as the *SaclayMocks* matter field without smoothing (Eqn. 3.13). I used  $\delta_m$  for comparison with the reconstructed Ly $\alpha$  contrast map  $\delta_{F_{\text{map}}}$  of the *SMocks-RSD* mock.

In this chapter, the binning of this field is  $d_{\text{cell}} = 2.19 h^{-1}\cdot\text{Mpc}$ . This scale is too small compared to the typical variation of the tomographic map controlled by  $L_{\parallel}$  and  $L_{\perp}$ . For this reason, I chose to apply a Gaussian smoothing with a length  $d_{\text{filter}} = 8 h^{-1}\cdot\text{Mpc}$  on the matter map. The value of this smoothing was chosen in order to optimize the correlation signal with the tomographic map. A representation of  $\delta_L$  and  $\delta_m^{\text{G}}$  fields is given in Fig. 6.6.

The large-scale matter density map is compared to the tomographic Ly $\alpha$  contrast map in Fig. 6.7. At high redshifts, a large fraction of the map is masked. Therefore, I restrict the comparison to the redshift range  $z = 2.1$  to 2.7.

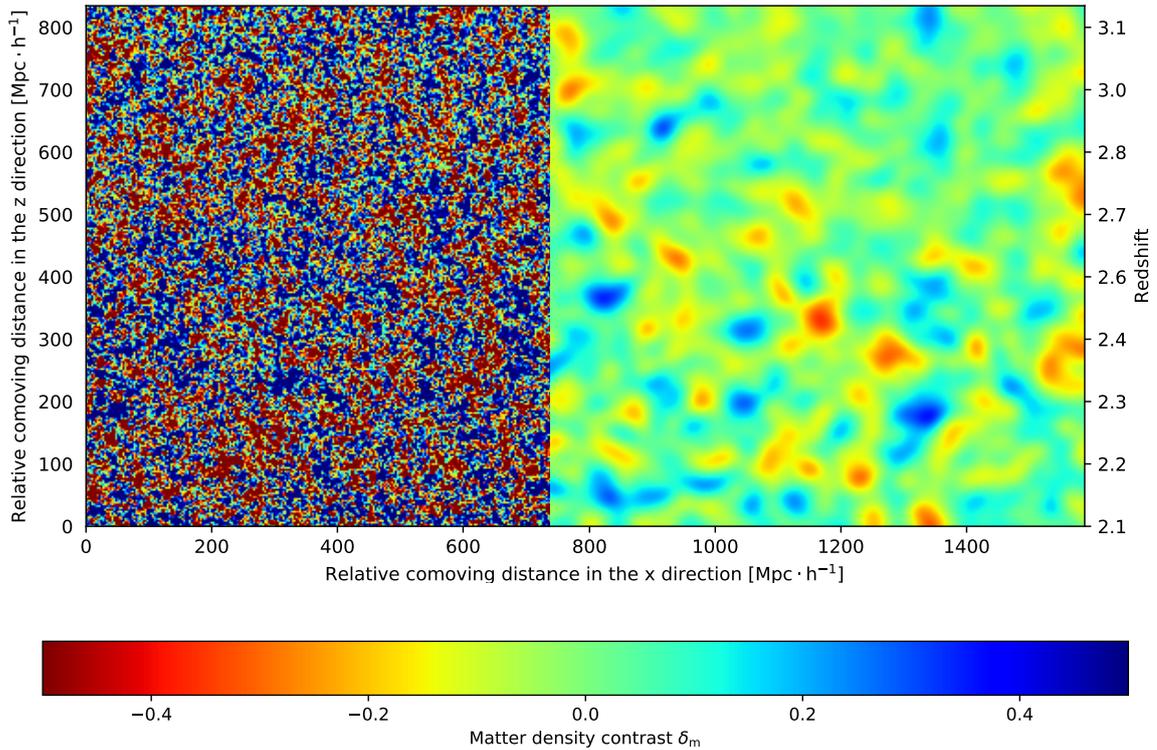


FIGURE 6.6 – Representation of underlying matter fields of the *SMocks-RSD* mock. The geometry of the slice shown corresponds to that of Fig. 6.5, which is one quarter of the Stripe 82 footprint. (left) Large-scale matter density field  $\delta_L$  defined by Eqn. 3.1. (right)  $\delta_m^G$  field with a Gaussian smoothing length  $d_{\text{filter}} = 8 h^{-1} \cdot \text{Mpc}$  over all directions.

The correlation coefficient between both fields is  $r = -0.34$ . This number provides an estimation of the ability of the tomographic map to trace the underlying cosmological field of matter fluctuations. This coefficient can be compared to simulation-based results in [8]. The authors tested several signal-to-noise ratios and line-of-sight separations and reported a correlation coefficient of 56% for the configuration closest to my dataset ( $\langle d_{\perp} \rangle = 12.65 h^{-1} \cdot \text{Mpc}$  and signal-to-noise ratio of 2). However, their correlation coefficient was computed relative to the initial Ly $\alpha$  contrasts, not to the underlying density fields.

It is also possible to compare this correlation factor to results from weak-lensing tomography of large-scale matter fluctuations. In particular, [42] used wide-field weak-lensing observations to construct a 3D tomographic map of matter fluctuations for  $z = 0.1 - 1$  over a  $167 \text{ deg}^2$  field. The comparison with matter density fluctuations as traced by galaxies yielded a correlation coefficient  $r \sim 10 \%$ .

As demonstrated in the simulations of [8], the correlation of the tomographic map with the underlying matter field would naturally improve with smaller  $\langle d_{\perp} \rangle$  and, to a smaller extent, larger signal-to-noise.

### 6.2.3 eBOSS data

I apply the algorithm detailed in Sec. 6.1.2 on eBOSS DR16 data presented in Sec. 6.1.1, with the same parameters and output format as those for the *SMocks-RSD* mock map presented in Sec. 6.2.1.

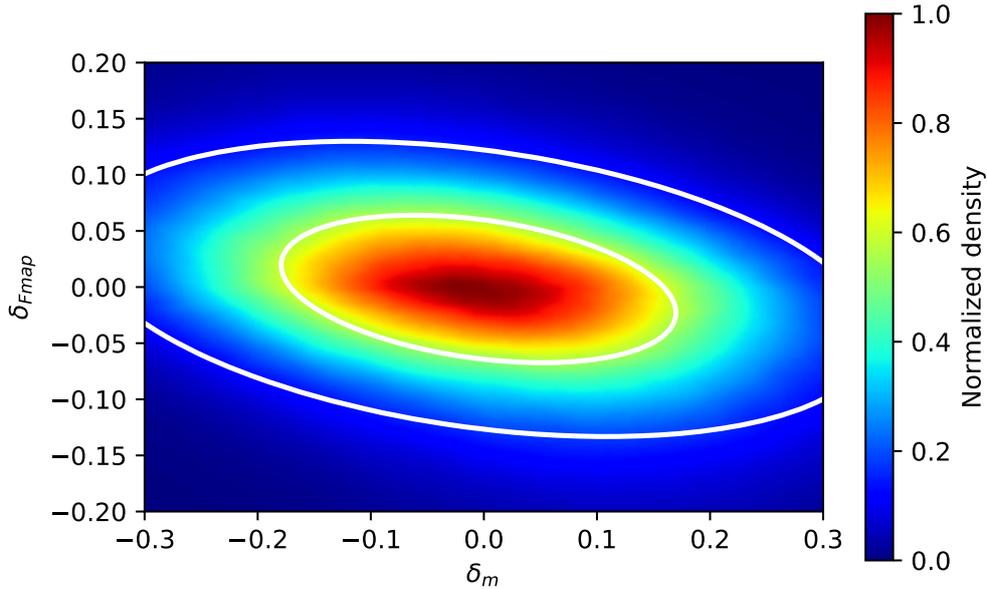


FIGURE 6.7 – Comparison between the reconstructed Ly $\alpha$  contrast from the *SMocks-RSD* mock and the large-scale matter density field  $\delta_L$ . The redshift range  $z = 2.1 - 2.7$  is considered. The map of large-scale matter density fluctuations was Gaussian smoothed over  $8 h^{-1}\cdot\text{Mpc}$ . White ellipses represent the  $1\sigma$  and  $2\sigma$  contours of a 2D Gaussian fit to the distribution.

The result of the tomographic reconstruction over the whole Stripe 82 is a map with  $(5856 \times 180 \times 834)$  pixels. Its total size in terms of  $(X, Y, Z)$  coordinates, in the right ascension, declination, redshift directions, is  $(6352 \times 180 \times 835) h^{-3}\cdot\text{Mpc}^3$ . The corresponding comoving volume, computed without parallel line-of-sight approximation, is  $0.94 h^{-3}\cdot\text{Gpc}^3$ .

This map is made publicly available on Zenodo<sup>4</sup>. Fig. 6.8 presents a constant-declination slice of this map over one quarter of the Stripe 82 field.

In addition, I worked in collaboration with a physicist of CEA/IRFU/DEDIP to create 3D representations of the tomographic map. Those visualizations were prepared with the Saclay Data visualization (SDvision) software [43] deployed in the IDL environment. This visualizations are also provided in an interactive web-based platform called Sketchfab<sup>5</sup>. The SDvision software allows for the interactive and immersive visualization of scalar and vector fields, clouds of points, and ancillary datasets such as the lines-of-sight and void catalogs. In my representations, the Ly $\alpha$  contrast is visualized using 3D iso-contour surfaces at negative and positive values to indicate overdense and underdense regions, respectively. Sketchfab is a web-based service that enables the sharing of 3D models. The user can rotate, pan, zoom in and out regions of interest, and benefit from GPU acceleration through the WebGL API, as well as immersive Virtual Reality capabilities in association with VR headsets. Examples and links to these 3D interactive animations are given in Fig. 6.9 and 6.10. All the 3D interactive animations are available online<sup>6</sup>.

4. <https://doi.org/10.5281/zenodo.3737781>

5. <https://sketchfab.com/>

6. <https://sketchfab.com/pomarede/collections/eboss-paper>

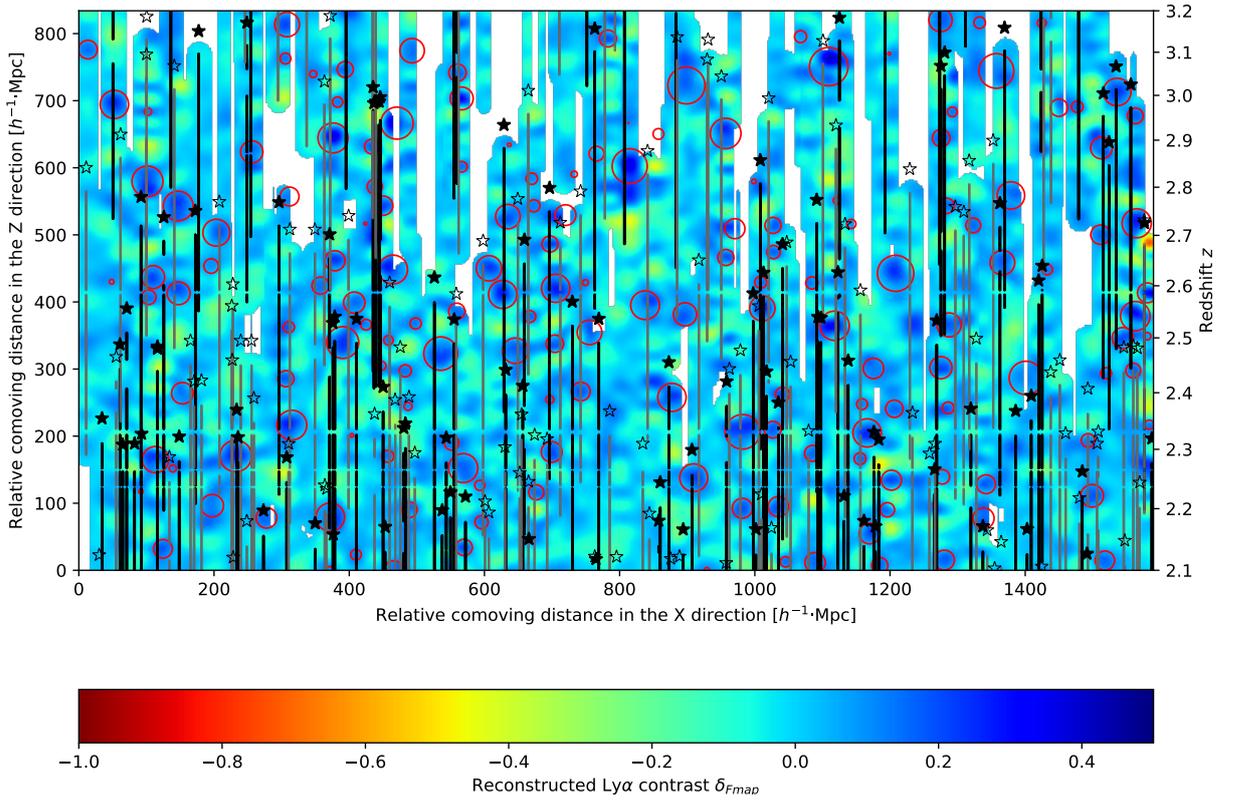


FIGURE 6.8 – Slice of the tomographic Ly $\alpha$  contrast map computed from the observed eBOSS data in the Stripe 82. The slice is at constant declination  $\delta_{J2000} = 0^\circ$ , and covers RA  $\in [1^\circ, 23^\circ]$ , i.e., roughly a quarter of the Stripe 82 field. The lines-of-sight, quasars, and voids are represented as in Fig. 6.5. All pictured quasars and line-of-sights (full and dotted) are located in a  $20 h^{-1}\cdot\text{Mpc}$ -thick slice. This thickness corresponds to  $0.28^\circ$  in the declination direction.

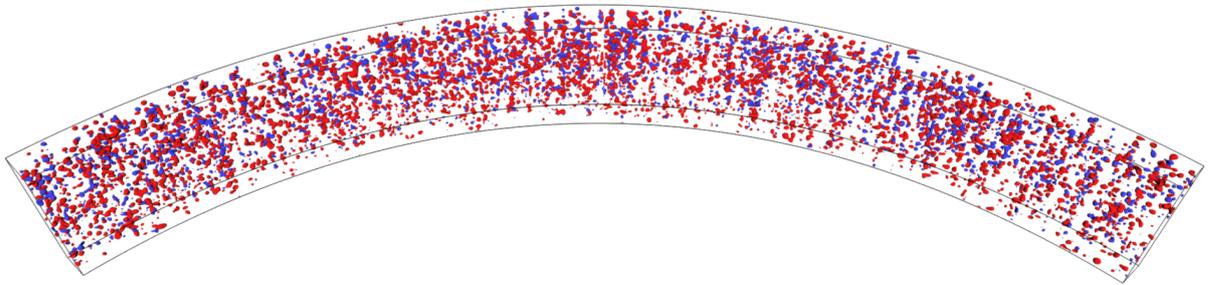


FIGURE 6.9 – 3D representation of the tomographic map in spherical coordinates over the entire Stripe 82. Red (blue) iso-surfaces of Ly $\alpha$  contrast equal to  $-0.2$  ( $0.2$ ) represent overdensities (underdensities). The corresponding 3D interactive representation is available on <https://skfb.ly/6RyGL>. An interactive representation of overdensities only is available on <https://skfb.ly/6RyGE>.

The 1D distribution of the reconstructed Ly $\alpha$  contrast in the tomographic map is presented and compared to that of the mock map in Fig. 6.11. A Gaussian distribution well approximates both histograms. However, I observe a higher standard deviation in the data than in the *SMocks-RSD* mock. This result may be related to the difference in noise distribution that is seen in Fig. 6.4.

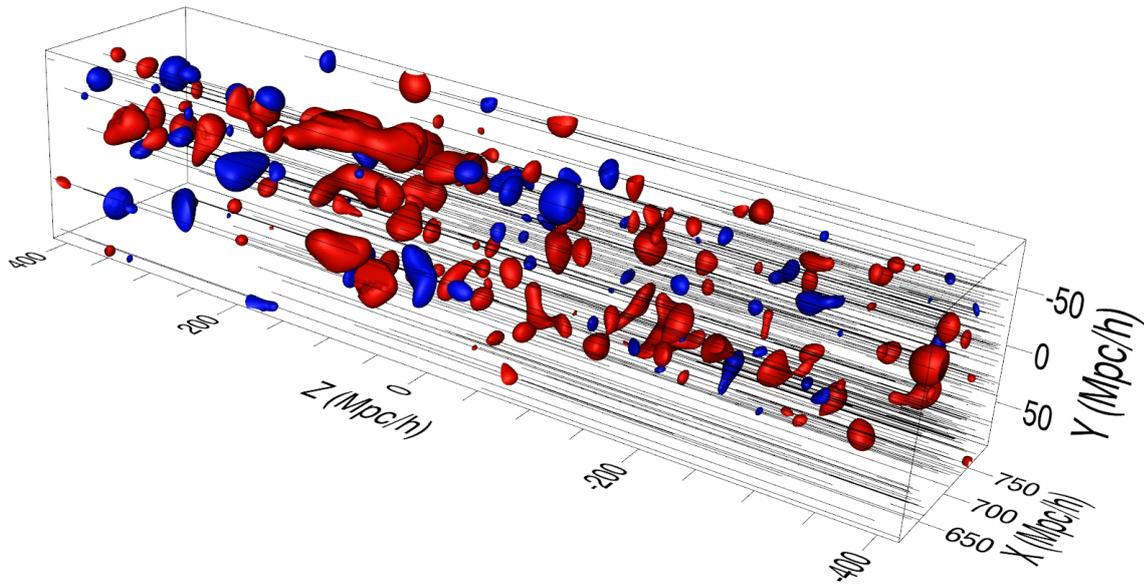


FIGURE 6.10 – 3D representation of a Stripe 82 portion which corresponds to  $2.5^\circ \times 2.5^\circ$  field in right ascension and declination. Red (blue) iso-surfaces of Ly $\alpha$  contrast equal to  $-0.2$  ( $0.2$ ) represent overdensities (underdensities). Lines-of-sight are represented in black. The corresponding 3D interactive representation is available on <https://skfb.ly/6RyGn>.

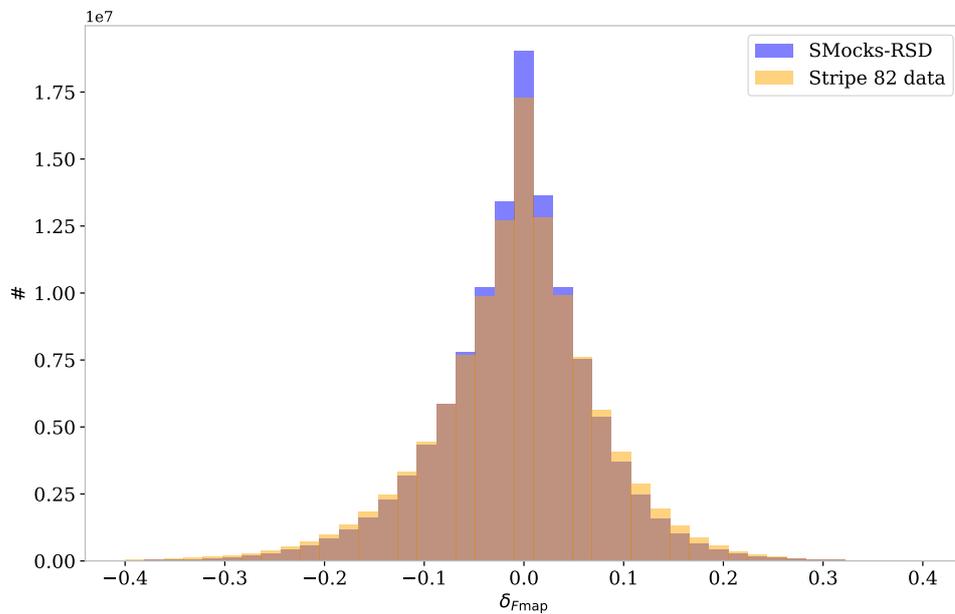


FIGURE 6.11 – Histogram of the Ly $\alpha$  contrast for tomographic map derived from the *SMocks-RSD* mock and for eBOSS DR16 data.

## 6.3 Applications

The map obtained from eBOSS data can be used for several applications similar to those realized from the CLAMATO map [4, 5]. Since the computed tomographic Ly $\alpha$  contrast map is anti-correlated with matter density fluctuations, an immediate application is to search for large under- and overdensities in the mapped volume. In this section, I present the detection of voids, candidate protoclusters, and the stacking of catalogs on the map performed in [19].

### 6.3.1 Void finder

**Voids** are underdensities in the cosmic web. They are the largest structures compared to nodes, walls, and filaments. Their typical size is approximately  $10 h^{-1}\cdot\text{Mpc}$  [44], but they have a wide distribution of dimensions. I can identify large voids with my tomographic map smoothed at scales of  $13 h^{-1}\cdot\text{Mpc}$ . The near-Gpc volume covered by Stripe 82 makes such a search particularly interesting. This tomographic map is an opportunity to study the feasibility of void cosmology at high redshift with the Ly $\alpha$  forest [13, 36].

#### Void finder algorithm

The search for voids in cosmology was until now done based on galaxy catalogs. In this case, the data are a set of points in a given volume. The void finder is then often based on a Voronoi tessellation, applied to create a density map on a regular grid. It is the case for example of the VIDE [11] algorithm. In the case of Ly $\alpha$  tomography, the map of Ly $\alpha$  contrast is already provided on a regular grid. The two most used methods to obtain underdense (or overdense) regions from a grid are the **spherical and watershed void finders**.

The spherical method selects all of the pixels that have a reconstructed Ly $\alpha$  contrast  $\delta_{F\text{map}}$  larger than a threshold  $\delta_{\text{th}}$ . At each selected point, a sphere is grown in radius until the mean Ly $\alpha$  contrast value  $\bar{\delta}_{F\text{map}}$  inside the sphere reaches an average limit  $\delta_{\text{av}}$ . Voids with a radius smaller than a certain value  $R_{\text{min}}$  are deleted.

In the watershed algorithm, all pixels whose  $\delta_{F\text{map}}$  is larger than the threshold  $\delta_{\text{ws}}$  are selected and then sorted into groups of neighboring pixels. For each void, the center is defined as the pixel with the largest  $\delta_{F\text{map}}$ . The void radius is defined by  $R = (3N_{\text{pix}}V_{\text{pix}}/4\pi)^{1/3}$  where  $N_{\text{pix}}$  is the number of pixels contributing to the void, and  $V_{\text{pix}}$  is the pixel volume. In addition, voids whose radius is smaller than a fixed parameter  $R_{\text{min}}$  are removed.

For all void detection methods, some voids overlap at the end of the procedure. This is particularly true for the spherical void finder, which looks for voids on very close pixels. There are several methods to remove overlapping voids. This step may seem trivial, but the void removal procedure may change the position of void centers.

A first method to remove overlapping voids is called **iteration**. For a given void in the catalog, I consider the group of overlapping voids and keep only the maximal-radius voids. When several voids have this radius in a group, the center of the remaining void is defined as the barycenter of these maximal-radius voids. This method defines the void center without position bias. Groups of overlapping voids are repeatedly created, and the procedure is applied until all remaining voids are separated. A second method called **cluster** consists of grouping the overlapping voids by clusters. The remaining voids are constituted in the same way as the iteration method inside each cluster.

In the `lelantos`<sup>1</sup> package, I have implemented the two methods of void detection and overlapping void deletion presented here. The detection algorithms are both easily parallelized. The watershed algorithm decomposes the map into sub-regions called chunks and searches in parallel for voids on each chunk. For the spherical method, parallelization is realized on the pixels selected to grow a sphere. The decomposition in chunks is also done for this method, as I found that it accelerates significantly the algorithm. Similar to the tomographic map creation explained in Sec. 6.1.2, overlap regions between chunks are used to remove edge effects. Once voids are detected in each region, the overlapping voids are suppressed and the sub-catalogs are merged. A second suppression of overlapping voids is performed on the whole catalog.

The output of the void finder algorithm is a catalog of voids characterized by a center, a radius, a mean contrast  $\bar{\delta}_{F_{\text{map}}}$  inside the void, and the  $\delta_{F_{\text{map}}}$  value at the center of the void. This void finder method is also implemented as an overdensity finder by searching at the lowest  $\delta_{F_{\text{map}}}$  values of the map.

Several quality cuts are then applied to the catalog to clean it up. These criteria are used to remove the voids located in areas where the tomographic reconstruction cannot operate without extrapolating. All these cuts are implemented in the `lelantos` package:

- (a) **Radius:** The voids are deleted according to their radius, with a maximum and minimum value. There is a slight difference between this cut and the  $R_{\text{min}}$  parameter explained earlier. Indeed,  $R_{\text{min}}$  intervenes in the process of overlapping void deletion.
- (b) **Edge of map:** Voids that have a part outside the map are removed.
- (c) **Distance to lines-of-sight:** For each void, I compute the distance  $d_{\text{los}}$  of its center to the nearest line-of-sight. The voids with a given percentage of volume  $dV$  located at more than a distance  $d_{\text{los}}$  are removed.
- (d) **Crossing by lines-of-sight:** Complementarily, I compute for each void the ratio of the length of crossing lines-of-sight  $d_{\text{cross}}$  to the radius of the void  $r_{\text{cross}} = d_{\text{cross}}/R_v$ . Requiring a minimum value for  $r_{\text{cross,min}}$  removes the "less crossed" voids.

### Void catalog for Stripe 82

I adopted a spherical algorithm to identify voids from the Stripe 82 tomographic map. This approach is more robust than the watershed algorithm on large scales as the watershed algorithm tends to provide complex void shapes [13].

For the spherical method, I determined the impact of its main hyper-parameters by varying them over a wide range. The parameter  $\delta_{\text{th}}$  mainly controls the depth of the voids while  $\delta_{\text{av}}$  controls the profile of the void, i.e., its size. The parameters  $\delta_{\text{av}}$  and  $\delta_{\text{th}}$  must be chosen simultaneously because of their correlated impacts on the void selection. If  $\delta_{\text{th}}$  is too low, the centers of the detected voids are offset relative to the maximal  $\delta_{F_{\text{map}}}$  inside the void. If  $\delta_{\text{th}}$  is too high, only the deepest voids are detected. For "reasonable" values of  $\delta_{\text{th}}$ , modifying  $\delta_{\text{av}}$  changes the void radii but does not move the center positions. These findings are in agreement with [13].

I adjusted the void detection parameters by applying the void finder to my mock tomographic map, in a similar approach to [13]. The minimal radius is fixed at  $R_{\text{min}} = 7 h^{-1} \cdot \text{Mpc}$ . Scanning over different algorithm parameters, I choose  $\delta_{\text{th}} = 0.14$  and  $\delta_{\text{av}} = 0.12$ . With this choice, the positions of large voids ( $\sim 25 h^{-1} \cdot \text{Mpc}$  radius) are stable when varying the parameters around the central values. For the removal of overlapping voids, I used the iteration method. For a spherical void finder, almost all voids are overlapping as they are detected on a large amount of

very close pixels. Therefore, the cluster suppression method is not suitable as it would remove most of the voids.

In the end of this procedure, I obtained a void catalog called *raw*. I applied quality cuts **(a)** to **(d)** as described previously, with parameters  $R_{\min} = 7 h^{-1}\cdot\text{Mpc}$ ,  $dV = 5 \%$ ,  $d_{\text{los}} = 20 h^{-1}\cdot\text{Mpc}$ , and  $r_{\text{cross},\min} = 1$ . The resulting void catalog is called *clean-tomo*.

After applying these criteria, the void filling factor, defined as the ratio of the total void volume over the total map volume, is 14 %, similar to the value of 15 % found by [13].

### Void properties

Voids from the *clean-tomo* catalog are represented by red circles on the slice view of Fig. 6.8 and by grey spheres on the 3D representation of Fig. 6.12. The catalog is available on Zenodo<sup>4</sup>. A total of 439 void candidates are detected in Stripe 82 with a radius larger than  $20 h^{-1}\cdot\text{Mpc}$ . The distribution of void radii is shown in Fig. 6.13, and compared to that of mock data. For both data and *SMocks-RSD* mock, a tail of distribution is observed: the profile is well fit by an exponential law  $dN/dR \sim \exp(-R/R_0)$ , with  $R_0 = 4.9 \pm 0.3 h^{-1}\cdot\text{Mpc}$  ( $5.6 \pm 0.3$ ) for data (mock). This feature is expected and comparable to the measured statistical properties of voids at lower redshift (see e.g. [45]).

In order to investigate the origin of my large-void distribution tail, I applied a **shuffling** procedure. It consists of keeping the lines-of-sight positions fixed by randomly shuffling the pixel values:  $\delta_F$  and the associated  $\sigma_{\delta_F}$  are shuffled simultaneously between lines-of-sight. In order to keep the redshift dependence of the  $\delta_F$  properties, I realized this shuffling by redshift slices of width  $\Delta z = 0.1$ . The idea of this data shuffle is to remove all physical correlation signal of the lines-of-sight while keeping the geometry of the data and its noise properties.

After constructing the map and finding voids on this shuffled data, the resulting void radius distribution is shown in Fig. 6.13 in dashed line. The high-radius tail observed in data is absent after shuffling: the number of voids with a radius larger than  $20 h^{-1}\cdot\text{Mpc}$  is 6 % of the one observed in non-shuffled data. On the other hand, the amount of void candidates with radii  $\lesssim 15 h^{-1}\cdot\text{Mpc}$  are almost identical in the case of true and shuffled data. This indicates that for these radii, noise contamination is important. However, the shuffling of data is not an exact null test and does not indicate that all voids with radii  $\lesssim 15 h^{-1}\cdot\text{Mpc}$  are necessarily false positive detections. As a comparison, in the case the galaxies, random Poisson point distribution gives very similar results concerning the distribution of voids [46, 47].

The results of my void finding procedure on the mock tomographic map is plotted in Fig. 6.5, where red circles represent identified voids. The distribution of void positions is relatively uniform over the map. The largest reconstructed void radius from mock is  $34 h^{-1}\cdot\text{Mpc}$  as can be seen in Fig. 6.13.

## 6.3.2 Catalog stacking

### 6.3.2.1 Void stack

Using the appropriate comoving coordinate system, I compute an average 3-dimensional void profile from the mock tomographic map by stacking the map around all identified void centers. This **stack** makes it possible to estimate the average shape of the reconstructed Ly $\alpha$  contrast within voids. As smaller voids are likely to be more contaminated by noise fluctuations (see previous discussion), I focus on voids with a reconstructed radius larger than  $L_{\perp} = 13 h^{-1}\cdot\text{Mpc}$ .

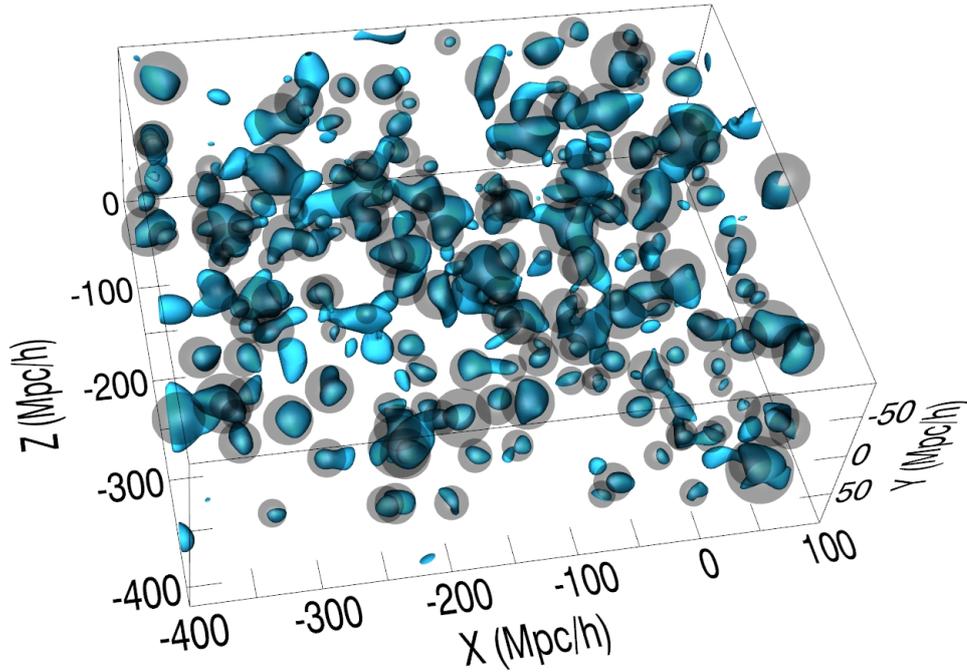


FIGURE 6.12 – 3D representation of a Stripe 82 portion which corresponds to  $7^\circ \times 2.5^\circ$  field in right ascension and declination. Blue iso-surfaces of Ly $\alpha$  contrast equal to 0.14 represent underdensities, at the same level as searched by the void finder ( $\delta_{\text{th}}$ ). Light grey spheroids represent identified voids. The corresponding 3D interactive representation is available on <https://skfb.ly/6RyGP>.

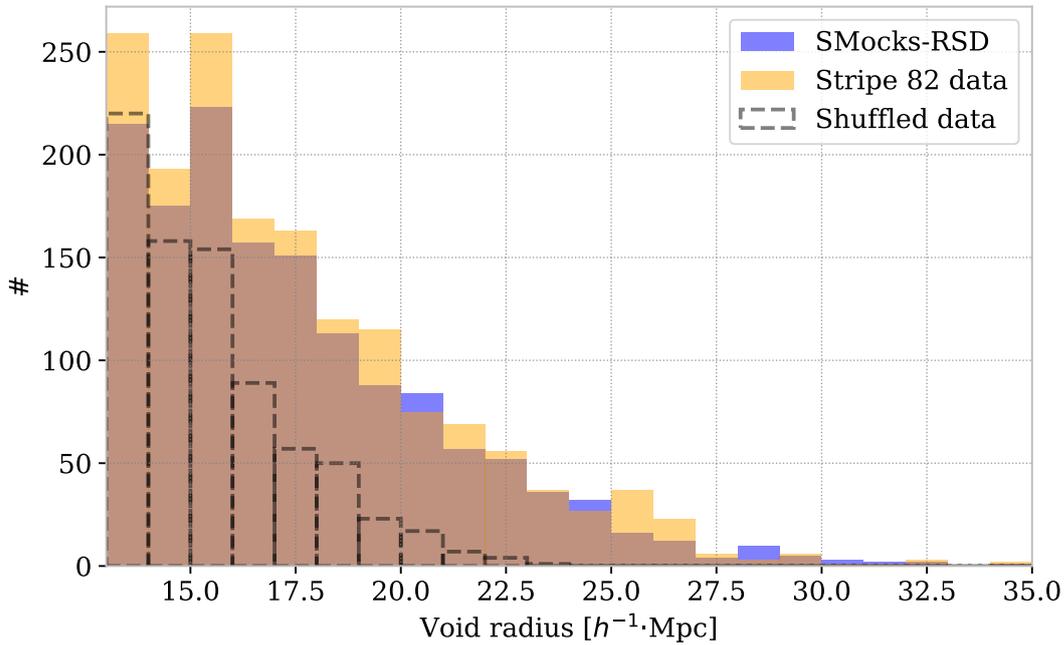


FIGURE 6.13 – Histogram of the *clean-tomo* void catalog (for radii larger than  $L_{\perp} = 13 h^{-1} \cdot \text{Mpc}$ ) detected on the *SMocks-RSD* mock and the eBOSS DR16 data. Dashed line represent void radii obtained after shuffling the eBOSS data.

The result is illustrated in Fig. 6.14. There are no fluctuations, given the large number of voids detected over a near-Gpc<sup>3</sup> volume.

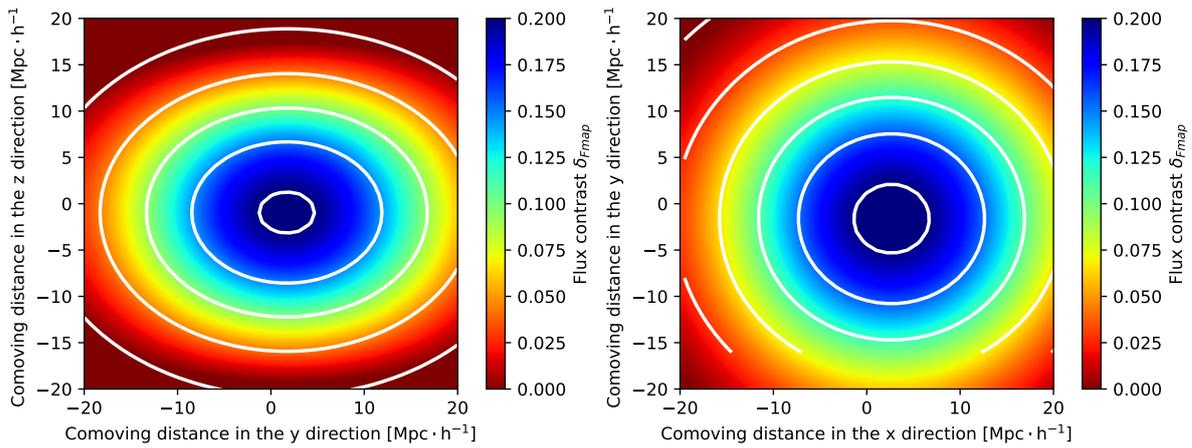


FIGURE 6.14 – Stacked profile of voids in the *clean-tomo* mock catalog, with radii larger than  $13 h^{-1} \cdot \text{Mpc}$ . The stack is performed overall the Stripe 82 field, with RSD effects included. White ellipses represent the contours of a Gaussian fit to the profile. The void stacks are not centered on zero. This issue was identified and corrected later in Chap. 7. Here, it does not affect the calculation of axis ratios. (left) Representation in the  $Y - Z$  plane. (right) Representation in the  $X - Y$  plane.

As shown in Fig. 6.14, the void profile has a circular shape in the  $X - Y$  plane, perpendicular to the lines-of-sight. However, it is anisotropic in the  $Y - Z$  plane as voids are more extended perpendicularly to the lines-of-sight. To quantify this effect, I fitted with the *lelantos* package the stacked void profile by a 2-dimensional Gaussian function. Axis ratios in the  $X - Y$ ,  $Y - Z$  and  $X - Z$  planes are defined by  $d_{ij} = \sigma_i / \sigma_j$  where  $\sigma_k$  is the standard deviation of the 2D Gaussian along the axis  $k$ . Statistical error bars on these  $d_{ij}$  values are computed with the Jackknife resampling method.

The measured axis ratios are reported on Tab. 6.1. There is a clear difference of axis ratio between the  $X - Y$  plane and the other planes for voids in the *SMocks-RSD* mock. There are three plausible causes to this measured 37 %-level anisotropy:

- **Coordinate system:** An error in the choice of the coordinate system, or equivalently in the underlying cosmological model, can generate an anisotropy in the average void profile - the Alcock-Paczynski effect [48]. Unless drastically departing from the coordinate system implied by existing constraints on cosmological parameters ([38]), such an effect cannot explain the observed axis ratio. I also checked that using the conventional  $(X, Y, Z)$  coordinate system, defined by Eqn. 6.8, does not modify the observed axis ratio.
- **Redshift-space distortions:** RSD are present in the  $\text{Ly}\alpha$  contrast, and taken into account in the *SMocks-RSD* mock as detailed in Sec. 6.2.1. To assess their impact on the average mock void profile, I compared the results to the *SMocks-noRSD* mock. Its axis ratio is  $d_{YZ} = 1.33 \pm 0.002$ , smaller than when RSD are included ( $d_{YZ} = 1.37 \pm 0.01$ ). This result demonstrates that RSD impact the observed void anisotropy but are not the dominant factor.
- **Data anisotropies:** While the Wiener filter algorithm I adopted is isotropic in itself ( $L_{\parallel} = L_{\perp}$ ), the input distribution of line-of-sight pixels is strongly anisotropic, with straight, parallel lines-of-sight. The tomographic map inherits this anisotropy. I realized a test by artificially sub-sampling the input line-of-sight of *SMocks-RSD* mock. In this case, the statistical isotropy of the corresponding pixels is increased, and I observed that the

Stack	Data set	$d_{yz}$	$d_{xz}$	$d_{xy}$
Voids	<i>SMocks-noRSD</i>	$1.33 \pm 0.002$	$1.31 \pm 0.002$	$0.97 \pm 0.001$
	<i>SMocks-RSD</i>	$1.37 \pm 0.01$	$1.36 \pm 0.01$	$0.98 \pm 0.005$
	Data	$1.24 \pm 0.001$	$1.22 \pm 0.003$	$0.97 \pm 0.003$
Quasars	<i>SMocks-noRSD</i>	$1.26 \pm 0.18$	$1.39 \pm 0.05$	$1.21 \pm 0.05$
	<i>SMocks-RSD</i>	$2.18 \pm 0.10$	$1.79 \pm 0.06$	$0.77 \pm 0.16$
	Data	$2.52 \pm 0.63$	$2.31 \pm 0.23$	$0.68 \pm 0.15$

TABLE 6.1 – Axis ratios for the void (illustrated in Fig. 6.14) and quasar (illustrated in Fig. 6.15) stacks. These ratios are compared between *SMocks-RSD*, *SMocks-noRSD* mocks, and Stripe 82 data. Voids from the *clean-tomo* catalog with radius larger than  $13 h^{-1}$  Mpc are used to perform the void stack.

mean void axis ratio is decreased. This behavior indicates that a significant fraction of the mean void axis ratio is due to this effect.

This stacking study has highlighted that RSD can be observed from the void profile. However, the signal is highly contaminated by the geometry of Ly $\alpha$  lines-of-sight. Another approach to this topic is presented in Chap. 7.

### 6.3.2.2 QSO stacking

In the Stripe 82 field, the only large and homogeneous sample of known objects with accurate (spectroscopic) redshifts  $z > 2.1$  are the quasars used as targets for BOSS/eBOSS Ly $\alpha$  forest measurements. Fig. 6.15 displays my Stripe 82 tomographic map stacked around quasars with redshift  $z = 2.1 - 3.2$ . There is a clear 3-dimensional signal extending over roughly  $20 h^{-1}$  Mpc. By stacking on random map positions, I quantified the significance of the central signal to be a  $5.2\sigma$  excess with respect to a random distribution. This result is a recast view of the small-separation part of the cross-correlation between quasar positions and the Ly $\alpha$  forest described in [39, 49, 50]. However, the statistical significance of the signal is smaller here since this study is limited to the Stripe 82 field.

Since quasars are, by definition, located at the end of lines-of-sight, there is a concern that the Wiener filtering algorithm itself biases the observed cross-correlation signal. However, the tomographic Ly $\alpha$  contrast is peaked right at the quasar position, within statistical uncertainties, whereas the lines-of-sight used for the reconstruction are located on average  $49 h^{-1}$  Mpc away, i.e., three times the width  $L_{\perp}$  of the filtering kernel. Therefore, if it exists, this bias is a minor effect.

A hint for anisotropy in the stacked signal around quasars is visible by eye in the  $Y - Z$  slice in Fig. 6.15. Using the same method as in Sec. 6.3.2.1, I computed the axis ratios of the stacked signal which is reported in Tab. 6.1, with statistical error bars estimated by Jackknife resampling.

I computed the same stack around quasars in the mock maps. Comparison between mocks and data in Tab. 6.1 shows that the anisotropy observed in the data is statistically sound and compatible with predictions from mocks. The origin of this anisotropy might be, as in the case of voids presented in Sec. 6.3.2.1, a combination of numerical effects in the map-making and RSD.

I also examined the correlation of the tomographic map with DLAs to try to reproduce the signal obtained by [51, 52]. I applied the same stacking procedure as was done for quasars,

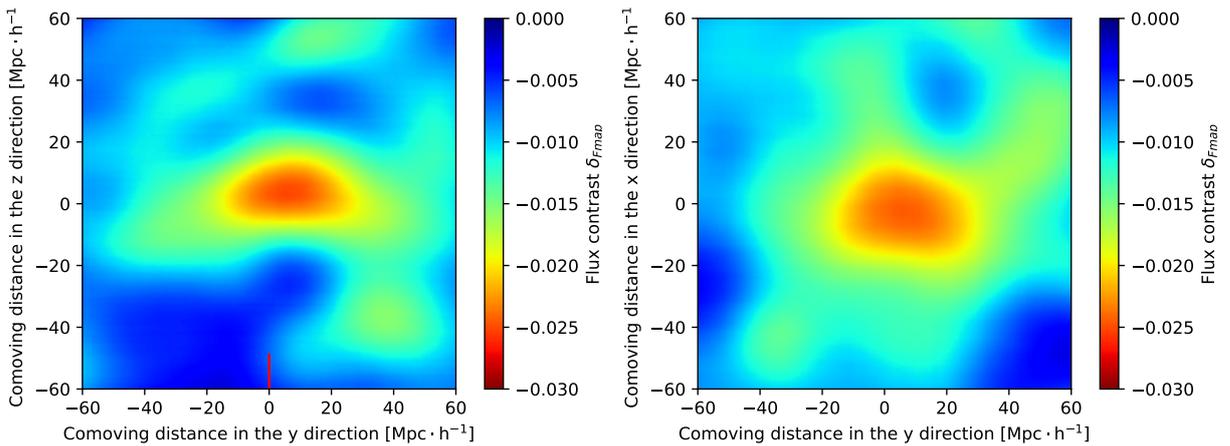


FIGURE 6.15 – Stacked tomographic Ly $\alpha$  contrast map centered around quasars over Stripe 82. (left) View in the  $Y - Z$  plane. The mean line-of-sight position with respect to the quasar is represented by a red line at the bottom of the image at  $Y = 0$ . (right) view in the  $X - Y$  plane.

centering on the positions of identified DLAs in the Stripe 82, with  $N_{\text{HI}} > 10^{20.3} \text{ cm}^{-2}$ . There is a  $2.6 \sigma$  deficit of tomographic Ly $\alpha$  contrast at the DLA positions. The signal is fainter than for quasars as there is an order of magnitude less detected DLAs.

I performed another stacking test with a photometric galaxy data set. The Hyper Suprime-Cam Subaru (HSC) program [53] detected galaxies at  $z > 2$  and estimated their photometric redshifts with machine learning methods. I tried to stack these galaxies on my tomographic map. This test did not yield conclusive results probably because of the limited galaxy statistics available (19,000) and because of the photometric redshift uncertainties that are diluting the signal.

### 6.3.3 Matter overdensities

I now present the result of blind searches for matter overdensities in the tomographic map.

Galaxy clusters are the largest bounded, overdense structures in the Universe, but they are formed late during cosmic evolution, at low redshift. **Protoclusters** are defined as structures that will, at some stage, collapse into a galaxy cluster. At  $z > 2$ , protoclusters are expected to be ubiquitous in the cosmic web and could have a characteristic radius of  $6 h^{-1} \cdot \text{Mpc}$  [54] for the largest ones. From the tomographic map produced in this study, it is clear that only overdensities with a typical scale larger than  $13 h^{-1} \cdot \text{Mpc}$  can be identified. Therefore, from my tomographic map, I may expect to identify only the largest protoclusters from  $z = 2.1$  to  $z = 3.2$  in the Stripe 82 field.

Based on the study of [36], I search for candidate high-redshift protoclusters or, more generally, large overdensities over the wide Stripe 82 field. To identify these overdensities, I search for large deficits in the Ly $\alpha$  contrast with the watershed algorithm presented in Sec. 6.3.1. My algorithm is similar to the one used by [18] to detect protoclusters in the CLAMATO survey.

In the protocluster search presented in [36], the CLAMATO tomographic map is smoothed beforehand at a  $4 h^{-1} \cdot \text{Mpc}$  scale, matching the dimension of searched protoclusters. In my case, I am limited by the  $13 h^{-1} \cdot \text{Mpc}$  smoothing length of my map, so this step is not required.

	1	2	3	4	5	6	7	8
RA $_{J2000}$ [deg]	-37.621	-27.442	3.1647	33.792	38.748	41.031	41.481	41.421
DEC $_{J2000}$ [deg]	0.583	-0.166	0.472	0.750	-1.166	-0.0276	-0.528	-0.0276
Redshift	2.734	2.818	2.525	2.435	2.377	3.170	2.111	2.764
Radius [ $h^{-1}$ ·Mpc]	15.13	10.54	13.88	10.68	9.081	7.623	16.20	12.77
Radius [arc min.]	12.56	8.75	11.52	8.869	7.539	6.329	13.45	10.60
Minimal $\delta_{F_{\text{map}}}$	-0.76	-0.44	-0.69	-0.63	-0.38	-0.40	-0.74	-0.54

TABLE 6.2 – Properties of selected protocluster candidates over all the volume of my Stripe 82 tomographic map. The minimal  $\delta_{F_{\text{map}}}$  value is located at the protocluster center.

The threshold  $\delta_{\text{ws}}$  impacts both the positions and the radii of overdensities. As in [13], I set  $\delta_{\text{ws}} = -3.5 \tilde{\sigma}_{\delta_{F_{\text{map}}}} = -0.273$ , where the standard deviation of the Ly $\alpha$  contrast of the map  $\tilde{\sigma}_{\delta_{F_{\text{map}}}}$  is computed by using a Gaussian fit of the  $\delta_{F_{\text{map}}}$  distribution.

I applied several quality cuts as defined in Sec. 6.3.1 on the resulting catalog to keep only the most significant protocluster candidates. I applied **(b)** and **(a)** cuts with  $R_{\text{min}} = 7 h^{-1}$ ·Mpc so that identified overdensities have a diameter larger than the smoothing length of the map. Residual HCD systems and metals can generate spurious overdensity detections, as pointed out by [16, 17]). To mitigate this effect, I applied **(d)** cut with  $r_{\text{cross,min}} = 6$ . This latter criterion implies in particular that each overdensity candidate is crossed by more than three lines-of-sight.

Applying my overdensity search procedure to the *SMocks-RSD* mock tomographic map results in the identification of nine candidates. Their mean radius is  $10.5 h^{-1}$ ·Mpc. The values of the underlying matter density fluctuations averaged within the overdensity radii are positive for all candidates, with a mean value  $\langle \delta_M \rangle = 0.4$ . Using the same procedure on the eBOSS data, I found eight protocluster candidates whose properties are listed in Tab. 6.2. Their average estimated radius is  $12 h^{-1}$ ·Mpc. As expected, they correspond to particularly extended and massive protocluster candidates. The number of candidate overdensities is compatible with the one found from the mock map.

I stress again that only very extended overdensities in the cosmic web can be detected using my tomographic map. Assessing in more detail the completeness and robustness of this selection and interpreting the protocluster properties would require the use of N-body simulations adapted to the identification of protocluster halos, which is beyond the scope of this thesis.

To try to attest the existence of these protoclusters, I performed some exploratory tests that have not been published in [19]. I tried to visually inspect the position of the eight protocluster candidates on other data sets. The only other type of available observations tracing matter at such high redshift is infrared data. First, I plotted the zoomed maps on a photometric infrared satellite survey called Wide-field Infrared Survey Explorer (WISE) [55, 56]. This test did not give convincing results - there were only three protoclusters located on the same field as WISE and no apparent overdensity. I also tried to look at the cosmic infrared background (CIB) map from Planck [57, 58]. The redshift kernel of this map is a bit higher than WISE. The protocluster positions were mostly located at the edges of the CIB map and there was no convincing signal.

## 6.4 DESI Lyman-break galaxies secondary target program

### LBG with DESI

Quasars are rare objects: their small density is a strong limiting factor for several cosmological applications. We have seen this in the case of Ly $\alpha$  tomography, and this is also true when using QSO as a tracer of matter (QSO $\times$ QSO or QSO $\times$ Ly $\alpha$  correlations), at  $z \geq 2$ . Lyman-break galaxies (LBGs) are the prime candidate to overcome this issue. They are high-redshift star-forming galaxies. Their spectra are almost entirely absorbed for wavelengths smaller than the Lyman limit,  $\lambda_\infty = 911.75 \text{ \AA}$ . Compared to QSO, they have a large sky density, but they are much fainter: it is more difficult to select them from photometric data, and more exposure time is required to measure their spectra.

LBGs are a major candidate for a future DESI-2 survey, to be done after the main survey, to improve the precision of cosmological measurements at  $z \geq 2$ . To assess the feasibility of LBG observations with DESI, a dedicated "secondary target" campaign was already carried out, in which I have been involved.

### Targeting

To identify candidate LBG from photometric data, we used the CLAUDS survey [59] which followed up the  $grz$  HSC deep fields performed on the CFHT. This survey contains some of the deepest  $u$ -band imaging to date (median depth of  $U = 27.1 \text{ AB}$ ) over areas relevant to cosmology, e.g.,  $\simeq 6 \text{ deg}^2$ . Of the CLAUDS fields, we chose the COSMOS field to have a common field with CLAMATO.

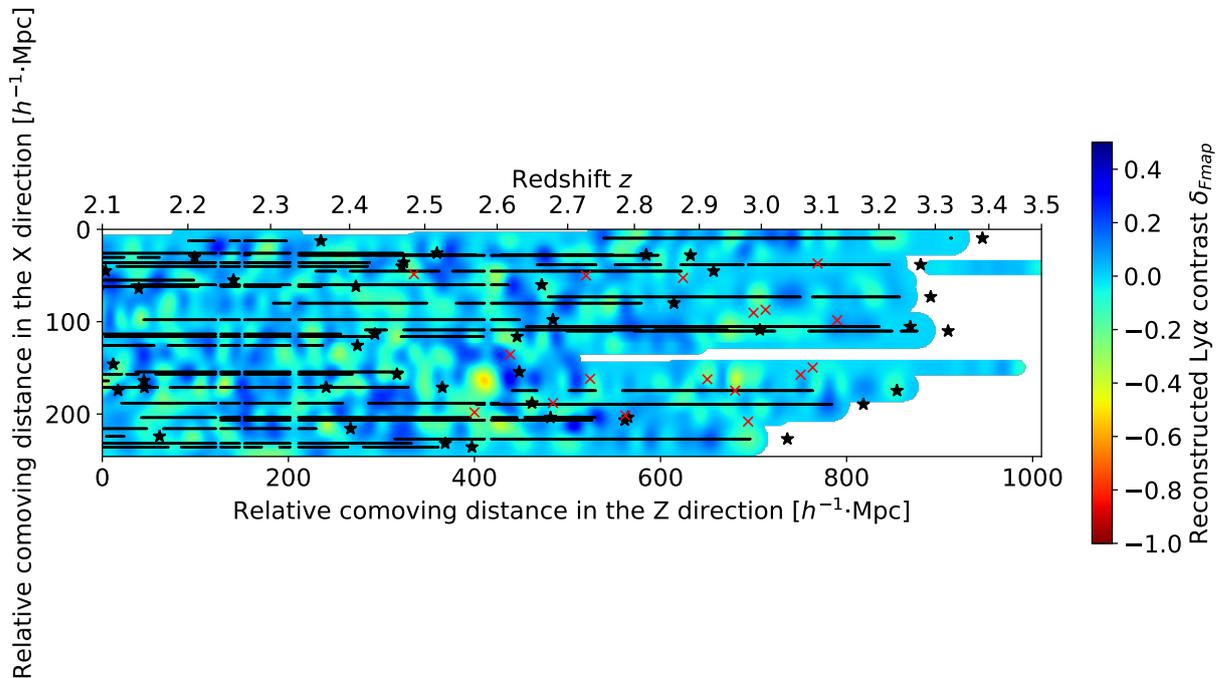


FIGURE 6.16 – Slice of the tomographic Ly $\alpha$  contrast map computed from the confirmed DESI QSO spectra in the COSMOS footprint. The same plotting conventions as Fig. 6.8 are applied here. Red crosses represent LBGs located at less than  $10 h^{-1} \text{ Mpc}$  of the slice.

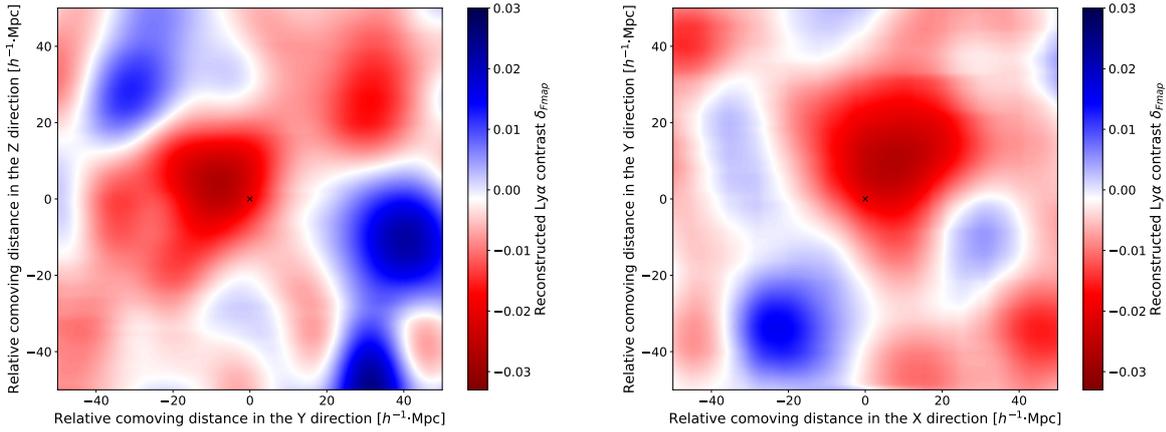


FIGURE 6.17 – Preliminary stacked DESI Ly $\alpha$  tomographic map centered around LBG positions. (left) View in the  $Y - Z$  plane. (right) view in the  $X - Y$  plane.

We selected objects with a bright continuum in order to be able to identify the Ly $\alpha$  forest in their spectra: QSOs and bright LBGs ( $r < 24.0$ ). For the QSOs, we used the standard DESI selection extended to  $r < 23.5$ , and with a color cut to mainly select QSOs with redshift  $z > 2.2$  (see Sec. 4.3.1). For the LBGs, we used a u-dropout method based on CLAUDS photometric catalogs [59]. The targeted LBGs cover redshifts  $2.0 < z < 3.5$ .

We obtained a total exposure of 5 hours over the COSMOS field. All the targets observed with this program were visually inspected, and we built a catalog containing 288 LBGs and 656 QSOs.

### First results

While we could, in principle, use both QSOs and LBGs to study Ly $\alpha$  absorption, at this stage, we use only QSOs. Indeed, the ability of DESI to use LBGs for Ly $\alpha$  forest studies is still under investigation, as it depends a lot on the achieved SNR in the spectra. Therefore, I created a Ly $\alpha$  tomographic map with the quasars observed in the COSMOS footprint. For that, I applied the Ly $\alpha$  tomography pipeline detailed in Sec. 6.1 with some modifications.

The spectra from 656 quasars has been collected on a footprint consisting in a single circular DESI tile of  $8 \text{ deg}^2$ , centered on  $(\text{RA}, \text{DEC}) = (150.0^\circ, 2.0^\circ)$ . The density of confirmed QSOs at  $z > 2.1$  is  $82 \text{ deg}^{-2}$ , larger than in eBOSS Stripe 82. The mean line-of-sight separation at  $z = 2.8$  is  $\langle d_\perp \rangle(z) \leq 10 h^{-1} \cdot \text{Mpc}$ , so we choose accordingly the reconstruction parameter  $L_\perp = 10 h^{-1} \cdot \text{Mpc}$ . The maximum redshift of the tomographic map is set to 3.5, in order to maximize the overlap with LBGs. A slice of the Ly $\alpha$  tomographic map is shown in Fig. 6.16. As before, this map is masked when the nearest line-of-sight is located at more than  $20 h^{-1} \cdot \text{Mpc}$ . The comoving dimensions of the map are  $(246 \times 261 \times 1010) h^{-3} \cdot \text{Mpc}^3$ , which corresponds to a total volume of  $0.053 h^{-3} \cdot \text{Gpc}^3$  when considering only redshifts  $2.1 < z < 3.2$ .

Here, we want to demonstrate that our observed LBGs can be used as matter tracers. For that purpose, I stacked their positions on the tomographic map: the result using 288 LBGs is shown in Fig. 6.17. We find that there is an excess of matter, as traced by the Ly $\alpha$  tomographic map, at the LBG position. By comparing to a random stacking, I estimate that the central overdensity has a  $3.5\sigma$  statistical significance. The presence of a signal with such a small statistic is extremely encouraging. This project is still in progress. On the one hand, we should be able to estimate

the bias of our selected LBG sample. On the other hand, the result supports the possibility of a DESI-2 survey using LBG, at least as matter tracers at  $z \geq 2$ .

### Summary and prospects

The large-scale Ly $\alpha$  tomography offers a unique opportunity to study the three-dimensional distribution of matter for scales of the order of  $10 h^{-1}\cdot\text{Mpc}$ . In this chapter, a map with the densest eBOSS field and one with the first DESI data were generated.

The improvement of the line-of-sight density and final footprint of DESI ( $60 \text{ deg}^{-2}$  over  $14000 \text{ deg}^2$ ) will allow the creation of a better map both in resolution and size. In the near future, the WEAVE-QSO [60] survey will observe a field of  $6000 \text{ deg}^2$  with a similar density to DESI. As forecasted in [61], the Ly $\alpha$  tomography will be able to be created in a denser field of WEAVE-QSO ( $110 \text{ deg}^{-2}$ ) and provide a map twice the size of Stripe 82.

Among the Ly $\alpha$  tomographic map applications, I focused in the next chapter on studying detected voids and the cross-correlation of their centers with the Ly $\alpha$  forest.

## Bibliography

- [1] P. Petitjean, *Clustering of absorption line systems*, *arXiv:astro-ph/9608115* (1996) .
- [2] K.-G. Lee, J. F. Hennawi, M. White, R. A. C. Croft and M. Ozbek, *Observational Requirements for Ly $\alpha$  Forest Tomographic Mapping of Large-scale Structure at  $z \sim 2$* , *The Astrophysical Journal* **788** (2014) 49.
- [3] K.-G. Lee, J. F. Hennawi, C. Stark, J. X. Prochaska, M. White, D. J. Schlegel et al., *Lyman-alpha Forest Tomography from Background Galaxies: The First Megaparsec-Resolution Large-Scale Structure Map at  $z > 2$* , *arXiv:1409.5632 [astro-ph]* (2014) .
- [4] K.-G. Lee, A. Krolewski, M. White, D. Schlegel, P. E. Nugent, J. F. Hennawi et al., *First Data Release of the COSMOS Ly  $\alpha$  Mapping and Tomography Observations: 3D Ly  $\alpha$  Forest Tomography at  $2.05 < z < 2.55$* , *The Astrophysical Journal Supplement Series* **237** (2018) 31.
- [5] B. Horowitz, K.-G. Lee, M. Ata, T. Müller, A. Krolewski, J. X. Prochaska et al., *Second Data Release of the COSMOS Lyman-alpha Mapping and Tomographic Observation: The First 3D Maps of the Large-Scale Cosmic Web at  $2.05 < z < 2.55$* , *arXiv:2109.09660 [astro-ph]* (2021) .
- [6] J. Japelj, C. Laigle, M. Puech, C. Pichon, H. Rahmani, Y. Dubois et al., *Simulating MOS science on the ELT: Ly $\alpha$  forest tomography*, *Astronomy & Astrophysics* **632** (2019) A94.
- [7] J. Cisewski, R. A. C. Croft, P. E. Freeman, C. R. Genovese, N. Khandai, M. Ozbek et al., *Non-parametric 3D map of the intergalactic medium using the Lyman-alpha forest*, *Monthly Notices of the Royal Astronomical Society* **440** (2014) 2599.
- [8] M. Ozbek, R. A. C. Croft and N. Khandai, *Large-scale 3D mapping of the intergalactic medium using the Lyman  $\alpha$  forest*, *Monthly Notices of the Royal Astronomical Society* **456** (2016) 3610.
- [9] M. R. Blanton, M. A. Bershad, B. Abolfathi, F. D. Albareti, C. A. Prieto, A. Almeida et al., *Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe*, *arXiv:1703.00052 [astro-ph]* (2017) .
- [10] R. Ahumada, C. A. Prieto, A. Almeida, F. Anders, S. F. Anderson, B. H. Andrews et al., *The Sixteenth Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra*, *arXiv:1912.02905 [astro-ph]* (2020) .
- [11] P. M. Sutter, G. Lavaux, N. Hamaus, A. Pisani, B. D. Wandelt, M. S. Warren et al., *VIDE: The Void IDentification and Examination toolkit*, *arXiv:1406.1191 [astro-ph]* (2014) .
- [12] M. Viel, J. M. Colberg and T.-S. Kim, *On the importance of high-redshift intergalactic voids*, *Monthly Notices of the Royal Astronomical Society* **386** (2008) 1285.

- [13] C. W. Stark, A. Font-Ribera, M. White and K.-G. Lee, *Finding high-redshift voids using Lyman  $\alpha$  forest tomography*, *Monthly Notices of the Royal Astronomical Society* **453** (2015) 4312.
- [14] A. Krolewski, K.-G. Lee, M. White, J. Hennawi, D. J. Schlegel, P. E. Nugent et al., *A Detection of  $z\sim 2.3$  Cosmic Voids from 3D Lyman- $\alpha$  Forest Tomography in the COSMOS Field*, *The Astrophysical Journal* **861** (2018) 60.
- [15] S. Mukae, M. Ouchi, Z. Cai, K.-G. Lee, J. X. Prochaska, S. Cantalupo et al., *3D Distribution Map of HI Gas and Galaxies Around an Enormous Ly $\alpha$  Nebula and Three QSOs at  $z=2.3$  Revealed by the HI Tomographic Mapping Technique*, *The Astrophysical Journal* **896** (2020) 45.
- [16] Z. Cai, X. Fan, S. Peirani, F. Bian, B. Frye, I. McGreer et al., *Mapping the Most Massive Overdensity Through Hydrogen (MAMMOTH) I: Methodology*, *The Astrophysical Journal* **833** (2016) 135.
- [17] Z. Cai, X. Fan, F. Bian, A. Zabludoff, Y. Yang, J. X. Prochaska et al., *Mapping the Most Massive Overdensities through Hydrogen (MAMMOTH). II. Discovery of the Extremely Massive Overdensity BOSS1441 at  $z = 2.32$* , *The Astrophysical Journal* **839** (2017) 131.
- [18] K.-G. Lee, J. F. Hennawi, M. White, J. X. Prochaska, A. Font-Ribera, D. J. Schlegel et al., *Shadow of a Colossus: A  $z=2.45$  Galaxy Protocluster Detected in 3D Ly- $\alpha$  Forest Tomographic Mapping of the COSMOS Field*, *The Astrophysical Journal* **817** (2016) 160.
- [19] C. Ravoux, E. Armengaud, M. Walther, T. Etourneau, D. Pomarède, N. Palanque-Delabrouille et al., *A tomographic map of the large-scale matter distribution using the eBOSS Stripe 82 Ly- $\alpha$  forest*, *Journal of Cosmology and Astroparticle Physics* **2020** (2020) 010.
- [20] N. Palanque-Delabrouille, C. Yèche, A. D. Myers, P. Petitjean, N. P. Ross, E. Sheldon et al., *Variability selected high-redshift quasars on SDSS Stripe 82*, *Astronomy & Astrophysics* **530** (2011) A122.
- [21] A. D. Myers, N. Palanque-Delabrouille, A. Prakash, I. Pâris, C. Yèche, K. S. Dawson et al., *The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Quasar Target Selection*, *The Astrophysical Journal Supplement Series* **221** (2015) 27.
- [22] B. W. Lyke, A. N. Higley, J. N. McLane, D. P. Schurhammer, A. D. Myers, A. J. Ross et al., *The Sloan Digital Sky Survey Quasar Catalog: Sixteenth Data Release*, *The Astrophysical Journal Supplement Series* **250** (2020) 8.
- [23] D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. A. Prieto, S. F. Anderson et al., *SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way Galaxy, and Extra-Solar Planetary Systems*, *The Astronomical Journal* **142** (2011) 72.
- [24] K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, S. Bailey et al., *The Baryon Oscillation Spectroscopic Survey of SDSS-III*, *The Astronomical Journal* **145** (2013) 10.

- [25] K. S. Dawson, J.-P. Kneib, W. J. Percival, S. Alam, F. D. Albareti, S. F. Anderson et al., *The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data*, *The Astronomical Journal* **151** (2016) 44.
- [26] S. Chabanier, T. Etourneau, J.-M. L. Goff, J. Rich, J. Stermer, B. Abolfathi et al., *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: The Damped Lyman- $\alpha$  systems Catalog*, *arXiv:2107.09612 [astro-ph, physics:physics]* (2021) .
- [27] D. Parks, J. X. Prochaska, S. Dong and Z. Cai, *Deep Learning of Quasar Spectra to Discover and Characterize Damped Ly $\alpha$  Systems*, *arXiv:1709.04962 [astro-ph]* (2017) .
- [28] S. Chabanier, N. Palanque-Delabrouille, C. Yèche, J.-M. L. Goff, E. Armengaud, J. Bautista et al., *The one-dimensional power spectrum from the SDSS DR14 Ly $\alpha$  forests*, *Journal of Cosmology and Astroparticle Physics* **2019** (2019) 017.
- [29] C. Pichon, J. Vergely, E. Rollinde, S. Colombi and P. Petitjean, *Inversion of the Lyman  $\alpha$  forest: three-dimensional investigation of the intergalactic medium*, *Monthly Notices of the Royal Astronomical Society* **326** (2001) 597.
- [30] S. Caucci, S. Colombi, C. Pichon, E. Rollinde, P. Petitjean and T. Sousbie, *Recovering the topology of the intergalactic medium at  $z \sim 2$* , *Monthly Notices of the Royal Astronomical Society* **386** (2008) 211.
- [31] N. Porqueres, O. Hahn, J. Jasche and G. Lavaux, *A hierarchical field-level inference approach to reconstruction from sparse Lyman- $\alpha$  forest data*, *arXiv:2005.12928 [astro-ph]* (2020) .
- [32] B. Horowitz, K.-G. Lee, M. White, A. Krolewski and M. Ata, *TARDIS Paper I: A Constrained Reconstruction Approach to Modeling the  $z \sim 2.5$  Cosmic Web Probed by Lyman-alpha Forest Tomography*, *The Astrophysical Journal* **887** (2019) 61.
- [33] B. Horowitz, B. Zhang, K.-G. Lee and R. Kooistra, *TARDIS Paper II: Synergistic Density Reconstruction from Lyman-alpha Forest and Spectroscopic Galaxy Surveys with Applications to Protoclusters and the Cosmic Web*, *The Astrophysical Journal* **906** (2021) 110.
- [34] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*. MIT Press, Cambridge, MA, USA, Aug., 1949.
- [35] W. H. Press, *Numerical recipes in C: The art of scientific computing, second edition*, *Endeavour* **17** (1993) 201.
- [36] C. W. Stark, M. White, K.-G. Lee and J. F. Hennawi, *Protocluster discovery in tomographic Ly  $\alpha$  forest flux maps*, *Monthly Notices of the Royal Astronomical Society* **453** (2015) 311.
- [37] Y. Saad, *Iterative Methods for Sparse Linear Systems*, Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics, Jan., 2003, 10.1137/1.9780898718003.
- [38] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astronomy & Astrophysics* **641** (2020) A6.

- [39] H. d. M. d. Bourboux, J. Rich, A. Font-Ribera, V. d. S. Agathe, J. Farr, T. Etourneau et al., *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Baryon acoustic oscillations with Lyman- $\alpha$  forests*, *The Astrophysical Journal* **901** (2020) 153.
- [40] A. González-Morales et al., *DESI Lyman-alpha synthetic spectra*, in preparation (2022).
- [41] H. du Mas des Bourboux, J. Rich, A. Font-Ribera, V. de Sainte Agathe, J. Farr, T. Etourneau et al., *picca: Package for Igm Cosmological-Correlations Analyses*, *Astrophysics Source Code Library* (2021) ascl:2106.018.
- [42] M. Oguri, S. Miyazaki, C. Hikage, R. Mandelbaum, Y. Utsumi, H. Miyatake et al., *Two- and three-dimensional wide-field weak lensing mass maps from the Hyper Suprime-Cam Subaru Strategic Program S16A data*, *Publications of the Astronomical Society of Japan* **70** (2018) .
- [43] D. Pomarede, H. M. Courtois, Y. Hoffman and R. B. Tully, *Cosmography and Data Visualization*, *Publications of the Astronomical Society of the Pacific* **129** (2017) 058002.
- [44] G. Lavaux and B. D. Wandelt, *Precision cosmography with stacked voids*, *The Astrophysical Journal* **754** (2012) 109.
- [45] P. M. Sutter, G. Lavaux, B. D. Wandelt, D. H. Weinberg, M. S. Warren and A. Pisani, *Voids in the SDSS DR9: observations, simulations, and the impact of the survey mask*, *Monthly Notices of the Royal Astronomical Society* **442** (2014) 3127.
- [46] S. Nadathur and S. Hotchkiss, *The nature of voids - I. Watershed void finders and their connection with theoretical models*, *Monthly Notices of the Royal Astronomical Society* **454** (2015) 2228.
- [47] S. Nadathur and S. Hotchkiss, *The nature of voids: II. Tracing underdensities with biased galaxies*, *Monthly Notices of the Royal Astronomical Society* **454** (2015) 889.
- [48] C. Alcock and B. Paczyński, *An evolution free test for non-zero cosmological constant*, *Nature* **281** (1979) 358.
- [49] H. d. M. d. Bourboux, J.-M. L. Goff, M. Blomqvist, N. G. Busca, J. Guy, J. Rich et al., *Baryon acoustic oscillations from the complete SDSS-III Ly $\alpha$ -quasar cross-correlation function at  $z=2.4$* , *Astronomy & Astrophysics* **608** (2017) A130.
- [50] M. Blomqvist, H. d. M. d. Bourboux, N. G. Busca, V. d. S. Agathe, J. Rich, C. Balland et al., *Baryon acoustic oscillations from the cross-correlation of Ly $\alpha$  absorption and quasars in eBOSS DR14*, *Astronomy & Astrophysics* **629** (2019) A86.
- [51] A. Font-Ribera, J. Miralda-Escudé, E. Arnau, B. Carithers, K.-G. Lee, P. Noterdaeme et al., *The large-scale cross-correlation of Damped Lyman Alpha Systems with the Lyman Alpha Forest: First Measurements from BOSS*, *Journal of Cosmology and Astroparticle Physics* **2012** (2012) 059.
- [52] I. Pérez-Ràfols, A. Font-Ribera, J. Miralda-Escudé, M. Blomqvist, S. Bird, N. Busca et al., *The SDSS-DR12 large-scale cross-correlation of Damped Lyman Alpha Systems with the Lyman Alpha Forest*, *Monthly Notices of the Royal Astronomical Society* **473** (2018) 3019.

- [53] H. Aihara, Y. AlSayyad, M. Ando, R. Armstrong, J. Bosch, E. Egami et al., *Second data release of the Hyper Suprime-Cam Subaru Strategic Program*, *Publications of the Astronomical Society of Japan* **71** (2019) 114.
- [54] R. A. Overzier, *The realm of the galaxy protoclusters*, *The Astronomy and Astrophysics Review* **24** (2016) 14.
- [55] T. H. Jarrett, M. Cohen, F. Masci, E. Wright, D. Stern, D. Benford et al., *The Spitzer-WISE Survey of the Ecliptic Poles*, *The Astrophysical Journal* **735** (2011) 112.
- [56] E. F. Schlafly, A. M. Meisner and G. M. Green, *The unWISE Catalog: Two Billion Infrared Sources from Five Years of WISE Imaging*, *The Astrophysical Journal Supplement Series* **240** (2019) 30.
- [57] S. J. Schmidt, B. Ménard, R. Scranton, C. B. Morrison, M. Rahman and A. M. Hopkins, *Inferring the redshift distribution of the cosmic infrared background*, *Monthly Notices of the Royal Astronomical Society* **446** (2015) 2696.
- [58] D. Lenz, O. Doré and G. Lagache, *Large-scale Maps of the Cosmic Infrared Background from Planck*, *The Astrophysical Journal* **883** (2019) 75.
- [59] M. Sawicki, S. Arnouts, J. Huang, J. Coupon, A. Golob, S. Gwyn et al., *The CFHT Large Area U-band Deep Survey (CLAUDS)*, *Monthly Notices of the Royal Astronomical Society* (2019) stz2522.
- [60] M. M. Pieri, S. Bonoli, J. Chaves-Montero, I. Paris, M. Fumagalli, J. S. Bolton et al., *WEAVE-QSO: A Massive Intergalactic Medium Survey for the William Herschel Telescope*, *arXiv:1611.09388 [astro-ph]* (2016) .
- [61] K. Kraljic, C. Laigle, C. Pichon, S. Peirani, S. Codis, J. Shim et al., *Forecasts for WEAVE-QSO: 3D clustering and connectivity of critical points with Lyman- $\alpha$  tomography*, *arXiv:2201.02606 [astro-ph]* (2022) .

# 7

---

## The correlation between Lyman- $\alpha$ forest and void positions

---

*“Comprenant l’interdépendance, on comprend la vacuité,  
Comprenant la vacuité, on comprend l’interdépendance,  
Telle est la Voie du Milieu  
Qui échappe aux terrifiants abysses du nihilisme et du réalisme.”*

– Kelzang Gyatso, *Selected Works of the Dalai Lama VII, XVIIIe siècle*

---

**Contents**

<b>7.1 Lyman-<math>\alpha</math> void cross-correlation pipeline</b>	<b>234</b>
7.1.1 Pipeline description	234
7.1.2 Pipeline testing	238
<b>7.2 Lyman-<math>\alpha</math> void cross-correlation on mocks</b>	<b>243</b>
7.2.1 Mock description	243
7.2.2 Matter void cross-correlation on mocks	244
7.2.3 Impact of systematic effects on the Lyman- $\alpha$ $\times$ Void cross-correlation	247
<b>7.3 Modeling the Lyman-<math>\alpha</math><math>\times</math>Void cross-correlation</b>	<b>253</b>
7.3.1 Redshift space distortions in Lyman- $\alpha$ $\times$ Void	253
7.3.2 Tomographic effect	256
7.3.3 Other tests and model summary	260
<b>7.4 Lyman-<math>\alpha</math><math>\times</math>Void cross-correlation in SDSS</b>	<b>263</b>
7.4.1 Data measurement and comparison with mocks	263
7.4.2 Model fitting	265
<b>Bibliography</b>	<b>268</b>

---

LARGE under-dense regions in the Universe, known as voids, provide an appealing way to study the cosmological evolution of large-scale structures. While structure formation is mostly studied using matter over-densities, e.g. from galaxy clustering measurements [1], voids were also considered as a complementary probe, using in particular the Galaxy $\times$ Void cross-correlation at redshift  $z < 1.5$  [2, 3]. This observable yields a measurement of the void density profile, but also the velocity flow around voids by taking advantage of RSD [4, 5, 6].

A key advantage of using voids for RSD studies is that linear theory better describes velocity flows around voids than around halos. At the same time, voids can also be used to carry out an Alcock & Paczynski consistency test [7] by stacking voids and studying their profile [8, 9].

In this chapter, I study for the first time RSD around voids at redshift  $z > 2$ , i.e., an epoch not reachable with galaxies yet. To achieve this, I correlate the Ly $\alpha$  forest data used to create the tomographic map of Chap. 6, with voids detected from this map. This work constitutes the first measurement of the cross-correlation between Ly $\alpha$  forest and voids.

As pointed out in Chap. 2, the Ly $\alpha$  forest is complicated by many modeling issues at the scales considered, down to  $h^{-1}\cdot\text{Mpc}$  scales. The physics is non-linear, and observational systematics are challenging to model. An advantage of voids is that their characteristic dimensions are the largest of the cosmic web. Therefore, we expect them to be less sensitive to systematics.

This chapter is outlined as follows: Sec. 7.1 details the measurement pipeline for the Ly $\alpha$  $\times$ Void cross-correlation as well as associated tests I performed. In Sec. 7.2, the mocks used for assessing the cross-correlation measurement are described. In order to interpret the results of the Ly $\alpha$  $\times$ Void cross-correlation, several models are built in Sec. 7.3. Finally, Sec. 7.4 presents the measurement on data and the comparison with the mocks. Using the developed models, I get a measurement of an RSD-related parameter around voids in this section.

This chapter is a more detailed view of [10]. For this analysis, I developed the python package `lyavoid`<sup>1</sup>. All algorithms used in this chapter are documented in this public package.

---

1. <https://github.com/corentinravoux/lyavoid/>

## 7.1 Lyman- $\alpha$ void cross-correlation pipeline

This section presents how the Ly $\alpha$ ×Void cross-correlation is performed and the tests made to validate the pipeline.

### 7.1.1 Pipeline description

The measurement of Ly $\alpha$ ×Void cross-correlation is performed in this chapter on the mocks presented in Sec. 6.2.1, and on the data from the SDSS-IV eBOSS survey’s Data Release 16 presented in Sec. 6.1.1.

#### Ly $\alpha$ ×Void cross-correlation definition

The **two-point cross-correlation function between the Ly $\alpha$  forest and the void centers (Ly $\alpha$ ×Void)**, noted  $\xi_{v\alpha}$ , is according to the formalism of Sec. 1.3.1:

$$\xi_{v\alpha}(\vec{r}) = \langle \delta_v(\vec{x}') \delta_F(\vec{x}' + \vec{r}) \rangle_{\vec{x}'} . \quad (7.1)$$

Here, we only consider information on void positions to define  $\xi_{v\alpha}$ : voids are considered as Boolean tracers like galaxies, so that  $\delta_v$  can be expressed:

$$\delta_v(\vec{x}') = \begin{cases} 1, & \text{if } \vec{x}' = \vec{X}, \quad \forall \vec{X}, \\ 0, & \text{elsewhere,} \end{cases} \quad (7.2)$$

where  $\vec{X}$  is the position of a void center. Therefore, the Ly $\alpha$ ×Void cross-correlation can be seen as the mean Ly $\alpha$  contrast around voids:

$$\xi_{v\alpha}(\vec{r}) = \langle \delta_F(\vec{X} + \vec{r}) \rangle_{\vec{X}} . \quad (7.3)$$

Note that we could use a different expression for  $\delta_v$ , taking into account the matter density around individual voids for example. This would however add a degree of complexity in the analysis.

#### Input data

To carry out the Ly $\alpha$ ×Void measurement, I use voids detected from the tomographic map in Chap. 6 and correlate their centers with the Ly $\alpha$  contrast along observed lines-of-sight, which are employed as an input for map-making.

Unlike Ly $\alpha$  tomography in Chap. 6, I generated maps pixelated at one pixel per  $2h^{-1}$ ·Mpc, which means twice less resolved. I applied this change for computational reason and checked that void finding and cross-correlation measurements were not impacted. The voids on the tomographic maps are detected as in Chap. 6 using a spherical void finder with the same parameters, and suppressing the overlapping voids with the iteration method. However, as the cut on the number of crossing lines-of-sight suppresses a significant number of voids, I no longer use it in this chapter. Indeed, for the Ly $\alpha$ ×Void cross-correlation, it is crucial to balance the purity of the voids and their statistics. After applying quality cuts, the resulting void catalog is called *clean-xcorr*. The calculation of the Ly $\alpha$ ×Void cross-correlation is applied either on *SaclayMocks* or eBOSS data.

For mocks, using the underlying matter density field presented in Sec. 6.2.2, I also measured the cross-correlation between the matter density field and the voids detected on this field, which will be referred to as Matter $\times$ Void and noted  $\xi_{\text{vm}}$ . The Matter $\times$ Void cross-correlation is defined by replacing the Ly $\alpha$  contrast by the matter density contrasts in Eqn. 7.1. To perform this measurement, I correlate the void center with an extraction of matter contrast along lines-of-sight. This procedure is done with the `lyavoid`<sup>1</sup> package.

For voids detected on the matter density maps, the void finder parameters are adapted to each map. Indeed, the density contrast values depend a lot on the transformations applied to the considered map. The measurement on these matter density fields will be detailed in the Sec. 7.2.2. For these void catalogs called *clean-matter-xcorr*, only the voids at the edge of the map are removed.

### Cross-correlation computation

The calculation of the Ly $\alpha$  $\times$ Void cross-correlation is done in the same way as the cross-correlation between the QSOs and the Ly $\alpha$  forest (Ly $\alpha$  $\times$ QSO) performed in [11, 12, 13, 14, 15, 16] and mentioned in Sec. 2.3. The `picca` [17] software is used for this purpose.

For each pair (Ly $\alpha$  contrast  $i$ , void  $j$ ), the transverse and parallel separations are defined from their redshifts  $z_i$  and  $z_j$ , and their angular separation  $\Delta\theta_{ij}$  computed from their RA and DEC coordinates. The definition of these separations is given by

$$\begin{aligned} r_{\parallel,ij} &= [D_C(z_i) - D_C(z_j)] \cos\left(\frac{\Delta\theta_{ij}}{2}\right), \\ r_{\perp,ij} &= [D_M(z_i) + D_M(z_j)] \sin\left(\frac{\Delta\theta_{ij}}{2}\right). \end{aligned} \quad (7.4)$$

In this calculation, the  $D_C$  and  $D_M$  comoving distances are computed using the same fiducial cosmological parameters as used for Ly $\alpha$  tomography in Chap. 6.

In addition to the  $(r_{\parallel}, r_{\perp})$  coordinates, I also use  $(r, \mu)$  defined by  $(r = \sqrt{r_{\parallel}^2 + r_{\perp}^2}, \mu = r_{\parallel}/r)$ . Those are necessary for the multipole calculation detailed hereafter.

For both coordinate systems, `picca` [17] creates bins  $A = (r, \mu)$  (respectively  $A = (r_{\parallel}, r_{\perp})$ ) with a number of bins labelled  $(N_r, N_{\mu})$  (respectively  $(N_{\parallel}, N_{\perp})$ ). The center of a given bin  $A$  is defined by the following pairwise renormalization in each bin:

$$\begin{aligned} r_A &= \frac{\sum_{(i,j) \in A} w_i r_{ij}}{\sum_{(i,j) \in A} w_i}, \\ \mu_A &= \frac{\sum_{(i,j) \in A} w_i \mu_{ij}}{\sum_{(i,j) \in A} w_i}. \end{aligned} \quad (7.5)$$

Here, the weight  $w_i$  is associated to a Ly $\alpha$  contrast pixel  $i$ . Those weights are used to optimize the signal-to-noise ratio in the correlation estimator. They are set to the inverse variance of the measured contrast:

$$w_i = \sigma_{\delta_F}^{-2}(\lambda_i) \left( \frac{1 + z_i}{1 + 2.25} \right)^{\gamma_{\alpha}-1}, \quad (7.6)$$

where  $\sigma_{\delta_F}$  is the Ly $\alpha$  contrast noise defined by Eqn. 5.9. The exponent  $\gamma_{\alpha} = 2.9$  [18] takes into account the redshift dependence of the Ly $\alpha$  forest bias. For a given  $\delta_m$ , the Ly $\alpha$  contrast is

higher at larger redshift so that we overweight them with respect to lower redshift. The pairwise binning implies that the centers are defined where most of the pairs are located.

The quasar continuum fitting step, detailed in Sec. 5.1.2, biases the Ly $\alpha$  contrast, with a significant impact on associated correlation measurements called distortion. Following [14, 15, 16], I correct these distortions by applying a **projection** on the Ly $\alpha$  contrast of each forest such that:

$$\delta_{F,i} \rightarrow \sum_j \eta_{ij} \delta_{F,j}, \quad (7.7)$$

where the  $i$  index runs over the wavelength, and the projection coefficient  $\eta_{ij}$  is defined by

$$\eta_{ij} = \delta_{ij}^K - \frac{w_j}{\sum_k w_k} - \frac{w_j (\Lambda_i - \bar{\Lambda}) (\Lambda_j - \bar{\Lambda})}{\sum_k w_k (\Lambda_k - \bar{\Lambda})^2}, \quad (7.8)$$

where  $\delta_{ij}^K$  is the Kroeneker delta,  $\Lambda_i$  is the function defined by Eqn. 5.5 which is equal to  $\log(\lambda_i)$  in the case of SDSS data, and  $\bar{\Lambda}$  is the average of  $\Lambda$  over all the forest. After this projection, the average of  $\delta_F$  along a given line-of-sight is not zero. We correct for this by redefining

$$\delta_{F,i} \rightarrow \delta_{F,i} - \bar{\delta}_F, \quad (7.9)$$

where  $\bar{\delta}_F$  is the weighted average of  $\delta_F$  along each line-of-sight with the weights defined by Eqn. 7.6.

Using this renormalized Ly $\alpha$  contrast, the Ly $\alpha$   $\times$  Void cross-correlation estimator for bin  $A$  is then given by the following formula:

$$\xi_{v\alpha}(A) = \frac{\sum_{(i,j) \in A} w_i \delta_{F,i}}{\sum_{(i,j) \in A} w_i}. \quad (7.10)$$

All the calculation steps for Ly $\alpha$  contrasts and the cross-correlation are identical to those performed for the Ly $\alpha$   $\times$  QSO cross-correlation. One difference in our case is that I do not add weights associated to voids.

The covariance of the cross-correlation between two bins  $A$  and  $B$  is defined by

$$C_{AB} = \langle \xi_{v\alpha}(A) \xi_{v\alpha}(B) \rangle - \langle \xi_{v\alpha}(A) \rangle \langle \xi_{v\alpha}(B) \rangle. \quad (7.11)$$

In this work, I chose to estimate it by a subsampling variance in HEALPix [19] coordinates, in the case when data are used. For the mocks, the covariance is defined in the same way when only one mock is produced. However, it can also be calculated from the variance between realizations when several mocks are created with identical parameters. When the subsampling method is used, the covariance estimator is given by

$$C_{AB} = \frac{1}{W_A W_B} \sum_H W_A^H W_B^H \left[ \xi_{v\alpha}^H(A) \xi_{v\alpha}^H(B) - \xi_{v\alpha}(A) \xi_{v\alpha}(B) \right], \quad (7.12)$$

where  $\xi_{v\alpha}^H$  refers to the cross-correlation calculated in a single HEALPix pixel  $H$ , and  $W_A = \sum_{(i,j) \in A} w_i$ .

The output product of `picca` [17] is composed of three arrays  $r$ ,  $\mu$  and  $\xi_{\text{v}\alpha}$ , of size  $N_r \times N_\mu$ , defining the cross-correlation for each bin center. It also contains the covariance matrix  $C$  of size  $(N_r \times N_\mu, N_r \times N_\mu)$ . This output is expressed in the same way with bins in  $(r_{\parallel}, r_{\perp})$ .

The same algorithms are applied to the Matter $\times$ Void cross-correlation except that the input coordinates are Cartesian  $(X, Y, Z)$  instead of  $(\text{RA}, \text{DEC}, z)$ . I adapted the `picca` [17] routines to operate in Cartesian coordinates. In that case, the computation of transverse and parallel separations does not require to assume fiducial cosmological parameters. With  $(X_i, Y_i, Z_i)$  and  $(X_j, Y_j, Z_j)$  the respective Cartesian coordinates of a Ly $\alpha$  contrast pixel and a void, the separations are simply defined by

$$\begin{aligned} r_{\parallel,ij} &= Z_i - Z_j, \\ r_{\perp,ij} &= \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2}. \end{aligned} \quad (7.13)$$

The output product of Matter $\times$ Void cross-correlation has the same format as the Ly $\alpha$  $\times$ Void cross-correlation.

### Multipole calculation

As indicated by Eqn. 7.3, the Ly $\alpha$  $\times$ Void (respectively Matter $\times$ Void) cross-correlation traces the profile of Ly $\alpha$  (respectively matter density) contrast around the voids. In this chapter, we need to use mathematical objects to quantify the angular anisotropy of this function. In our context, the optimal basis for expressing these deformations is the Legendre polynomials  $P_\ell(\mu)$ . They express the anisotropies of a function which depends on  $r$  and  $\mu$  thanks to the **multipole decomposition**

$$\xi_{\text{v}\alpha,\ell}(r) = \int_{-1}^1 \xi_{\text{v}\alpha}(r, \mu) \left( \frac{1 + 2\ell}{2} \right) P_\ell(\mu) d\mu, \quad (7.14)$$

where  $\xi_{\text{v}\alpha,\ell}$  is the **multipole of order  $\ell$**  of the Ly $\alpha$  $\times$ Void cross-correlation function. If this function is of the type  $A + B\mu^2$ , then only the multipoles of order 0 (monopole) and 2 (quadrupole) are non-zero. This result will be used latter in the RSD analysis.

I compute the multipole expansion of the Ly $\alpha$  $\times$ Void cross-correlation expressed in bins  $A = (r, \mu)$ . As indicated in Eqn. 7.5, this binning is not necessarily regular due to pairwise binning. For each line along with the  $r$  array, I choose to associate its mean value to the output multipole pixel. Concerning the  $\mu$  binning, I include the pairwise variations in the integration.

Secondly, because of the binning, the values of  $\mu$  do not reach  $\pm 1$  as maximal and minimal values. However, the even Legendre polynomials reach their maxima precisely at  $\pm 1$ : Eq. 7.14 shows that neglecting  $\mu \sim \pm 1$  in the integral leads to miscalculation of the even multipoles. The solution adopted is to linearly extrapolate the cross-correlations to cover the full range  $\mu \in [-1, 1]$ .

I implemented several multipole integration techniques in the `lyavoid` package. First, I used fixed samples integration methods such as rectangle, trapezoidal, and Simpson's rules. These methods directly integrate the cross-correlation for each  $r$  lines using the implementations of `scipy` [20]. I also refined the  $\mu$  values for each  $r$  line before applying the fixed sample integration methods. It reduces the separation between the pixels and thus decreases the integration error.

Finally, I also used functional integration methods given by `scipy` [20] such as quadratic and Romberg integrations after linearly interpolating each line in  $r$ . These methods gave similar

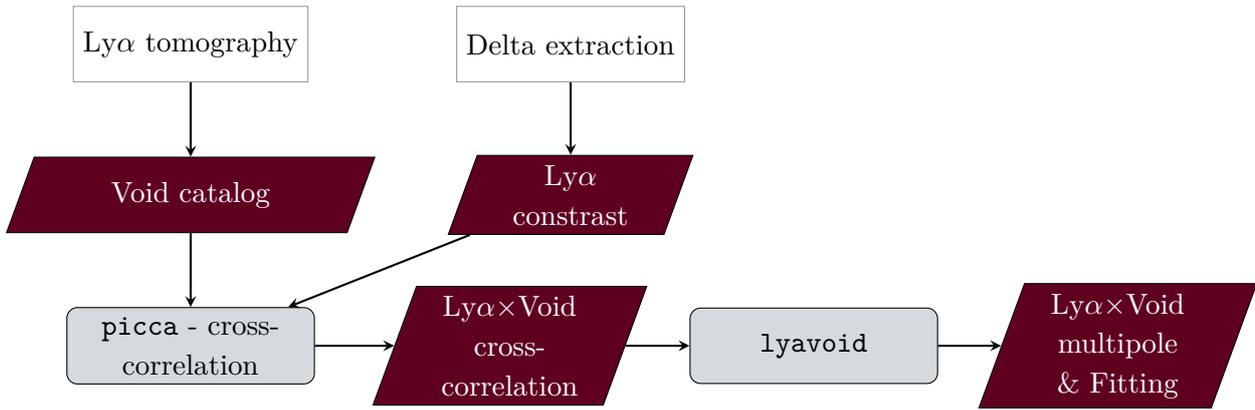


FIGURE 7.1 – Schematic representation of the Ly $\alpha$ ×Void cross-correlation pipeline. The steps of Ly $\alpha$  tomography and delta extraction correspond to those of the Ly $\alpha$  tomography pipeline represented in Fig. 6.3.

results to the rebinned fixed sample integration in tests of the Sec. 7.1.2. However, the functional integration methods increase the computational time.

To decide which integration method to use, I created an idealized cross-correlation with a regular grid and no dependence in  $\mu$  (perfectly round void profile). On this example, I selected the methods which, for a given binning  $N_\mu$ , minimized the integration error in the calculation of the multipoles. The Simpson’s sample integration and Romberg functional method gave the lowest residual multipole error for the same binning than data. To avoid introducing an additional interpolation step, Simpson’s rule integration after a thinner rebinning is chosen to calculate all multipoles.

A representation of the pipeline used to perform Ly $\alpha$ ×Void cross-correlation, the calculation of its multipole, and the fitting detailed in Sec. 7.4.2 is illustrated in Fig. 7.1. The Matter×Void cross-correlation is computed using the `picca` Cartesian cross-correlation. In this case, the matter contrast lines-of-sight are generated with `lyavoid`.

## 7.1.2 Pipeline testing

The goal of the tests performed in this section was to validate the ability of the pipeline in Fig. 7.1 to measure multipoles on fully controlled examples.

### Box tests

I generated examples containing ideal voids to test the void finder and the cross-correlation pipeline simultaneously. The ideal void profile used is the monopole of the mock Matter×Void cross-correlation, given in Sec. 7.2.2 (see Fig. 7.6). I created four boxes represented in Fig. 7.2 with different density field configurations:

- **Single void:** Box containing a single isotropic void which is used to test the void finder. This box allowed me to ensure that the center of the reconstructed void was not systematically displaced (for example, due to discretization effects). On the corresponding plot in Fig. 7.2, no overlapping void deletion is applied. This test certifies that the overlapping deletion method is adequate for finding the right void center. I verified explicitly that the input center position is the same as the void finder.

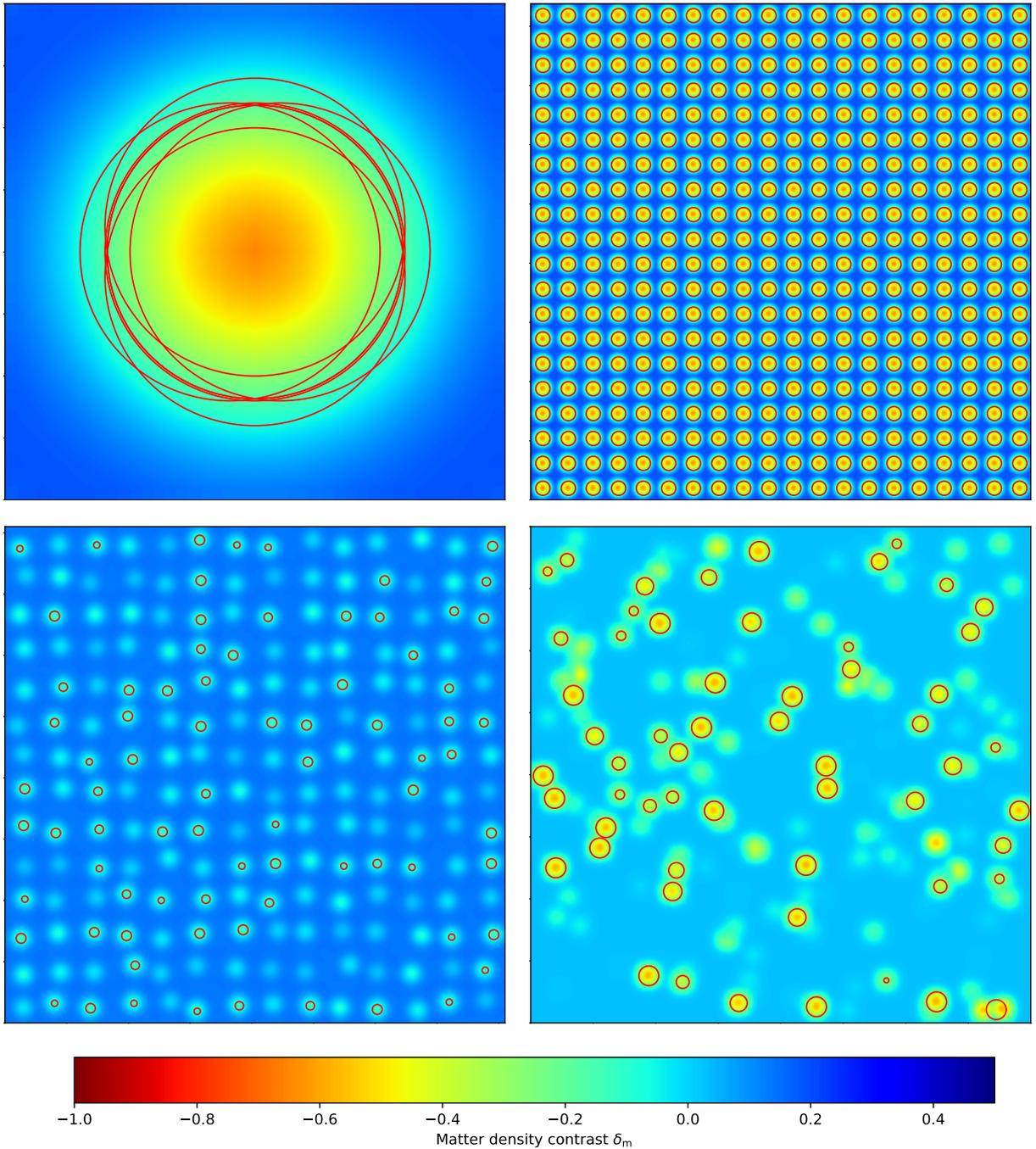


FIGURE 7.2 – Slices of the boxes used to test the algorithms developed in the Ly $\alpha$  tomography and Ly $\alpha$ ×Void cross-correlation pipelines. (top left) Single void. (top right) Regular. (bottom left) Regular random. (bottom right) Random. The void profile used for all these boxes originated from Matter×Void cross-correlation in Fig. 7.6. All the boxes are cubic, with a side size of  $20 h^{-1}\cdot\text{Mpc}$  for the top left panel, and  $400 h^{-1}\cdot\text{Mpc}$  for all the other panels.

- **Regular:** This is a box containing identical, regularly spaced voids. I used this box essentially to test the reconstruction of voids centers when the finder is multi-processed and the map is split into chunks during this task.
- **Regular random:** The voids in this box are defined from the Regular box by slightly moving the void center positions. Here, the shift is low enough to prevent the voids from

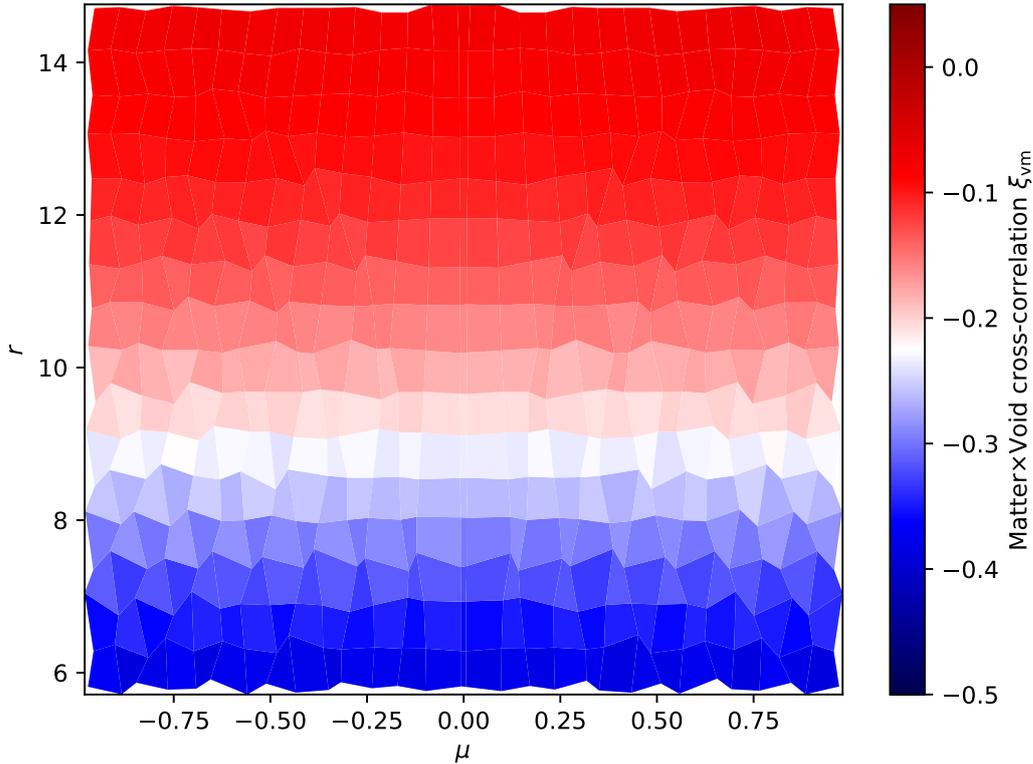


FIGURE 7.3 – Cross-correlation measurement with the Random box, measured in  $A = (r, \mu)$  space. The represented array is the one directly used to measure the multipoles, and it already appears that the cross-correlation is isotropic. The irregularity of the binning comes from the pairwise normalization given in Eqn. 7.5.

overlapping. I applied a Gaussian smoothing to have a smooth junction between the voids. Due to this smoothing, the depth of each void is modified and the only the deepest voids are detected by the algorithm.

- **Random:** In this last test, I randomly drew the position of a fixed number of single voids with identical profiles. I also applied a Gaussian smoothing. This box was used for the multipole calculation.

For all these examples, the parameters of the spherical void finder are adapted to the box smoothing. The parameter  $\delta_{av}$  mainly controls the radius of the void. Since I am mainly interested in retrieving the right center and not the radius, I only optimized the  $\delta_{th}$  parameter in order to find all voids on regular boxes.

In the case of the somehow realistic Random box, the Matter  $\times$  Void cross-correlation is shown in Fig. 7.3, in  $A = (r, \mu)$  space. For this calculation, the cross-correlation is computed for  $r \in [1, 15] h^{-1} \cdot \text{Mpc}$ ,  $\mu \in [-1, 1]$ , with a number of bins  $N_r = 25$  and  $N_\mu = 30$ . We also computed this cross-correlation in the  $A = (r_{\parallel}, r_{\perp})$  space, shown in Fig. 7.4 (left). For the two-dimensional plot, a binning such that  $r_{\parallel} \in [-20, 20] h^{-1} \cdot \text{Mpc}$ ,  $r_{\perp} \in [0, 20] h^{-1} \cdot \text{Mpc}$ ,  $N_{r_{\parallel}} = 40$ , and  $N_{r_{\perp}} = 20$  is used.

Comparing these two 2D representations, we see that the pairwise binning gives different results between the two  $A$  spaces. In particular, I do not show the correlation for  $r \lesssim 6 h^{-1} \cdot \text{Mpc}$  in  $A = (r, \mu)$  space because most of the bins in  $\mu$  are not filled for these  $r$  values. We interpret

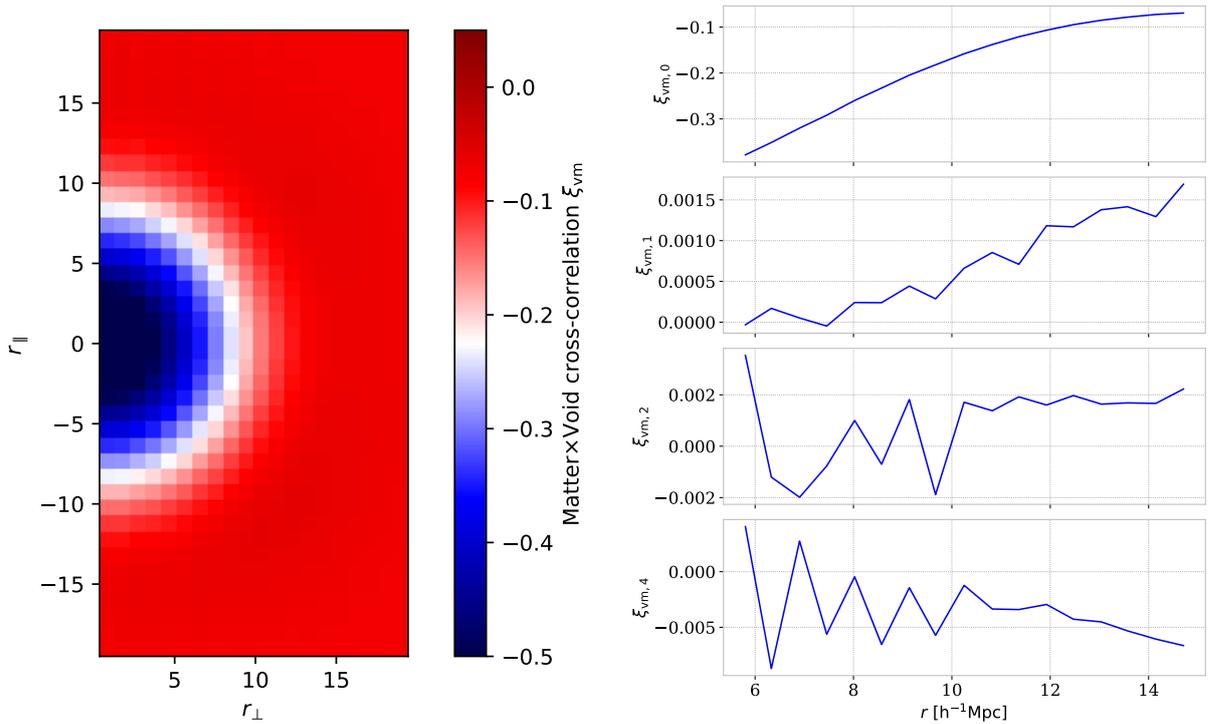


FIGURE 7.4 – (left) Cross-correlation 2D representation with the Random box in  $A = (r_{\parallel}, r_{\perp})$  binning. (right) Multipole of the associated cross-correlation reported in Fig. 7.3. The multipoles are computed using Simpson’s rule integration after rebinning in the  $\mu$  direction with  $N_{\mu} = 200$ .

the irregularity of the binning of Fig. 7.3 by the lack of statistics in this example. All the cross-correlations measured in the following sections do not present these irregularities.

We use the cross-correlation measured in the  $A = (r, \mu)$  space to calculate its multipoles, given in Fig. 7.4 (right). The monopole  $\xi_{\text{vm},0}$  gives the average shape of the void, and we find well in the figure the average input void profile. The dipole  $\xi_{\text{vm},1}$  signs an offset in the position of the void center. The quadrupole and hexadecapole ( $\ell = 2, 4$ ) characterize the anisotropies of the void profile (e.g., if the pipeline treats differently the  $X$ ,  $Y$ , and  $Z$ ). In the absence of numerical effect, and with an infinite statistics, we expect to have  $\xi_{\text{vm},\ell} = 0$  for  $\ell = 1, 2, 4$  in the case of perfectly spherical voids. In the Random box developed here, I obtain that the ratio  $|\xi_{\text{vm},\ell=1,2,4}|/|\xi_{\text{vm},0}|$  is less than 1% for all represented multipoles and separations ( $r$ ). I conclude that the cross-correlation pipeline I developed does not introduce numerical effects at a level larger than  $\sim 1\%$ . This is sufficient for the study performed in this chapter. In particular, the consistency of the dipole with zero validates the ability of the void finder algorithm to find the right void centers.

### Algorithm improvement

These tests on simplified boxes, and also on the mocks presented in the following section, allowed me to improve my void finder. Indeed, during a first iteration of tests, I found a quite large non-zero dipole for controlled examples. The efforts I made to find the origin of this dipole led to the following improvements to the cross-correlation pipeline:

- **Impact of parallelization:** The separation in different chunks, needed to make the code parallel, can lead to misalignments during merging. I had to improve the related bookkeeping to control this error.
- **Void merging:** Using the single void box represented in Fig. 7.2, I defined the center of the merged void as the barycenter of the largest-radius overlapping voids (as detailed in Sec. 6.3.1). This new definition helped to reduce the dipole.
- **Map coordinates:** An error on the conversion between position in pixel and distance in the map was biasing the void position. Again, solving this issue helped to obtain a null dipole in test boxes.
- **Multipole integration:** An error on the quadrupole at low  $r$  was also corrected. Without the extrapolation in  $\mu$  up to the  $\pm 1$  values detailed in the Sec. 7.1.1, the even multipoles were impacted at low  $r$  in first iteration of the tests.

## 7.2 Lyman- $\alpha$ void cross-correlation on mocks

Mocks used to interpret the Ly $\alpha$ ×Void cross-correlation were already presented in Sec. 3.1.1.

### 7.2.1 Mock description

As described in Sec. 3.1.1, **Sac1ayMocks** are used to understand the impact of various physical effects on the Ly $\alpha$  correlations, as well as systematics. In this chapter, I consider the following effects of interest, using the following acronyms:

- **RSD**: Redshift space distortions included in the calculation of the optical depth, with the  $c_{GP}$  term in Eqn. 3.9.
- **D**: Including DLA systems in the spectra thanks to the **quickquasars** software.
- **Dm**: Masking the abovementioned DLA systems, at the **picca** stage, after having used the DLA finder **dla\_cnn**.
- **N**: Modeling of the SDSS instrumental noise with **quickquasars**.
- **C**: Multiplying the Ly $\alpha$  transmission by the quasar continuum, using **quickquasars**.
- **CF**: Fitting of the quasar continuum at the **picca** Ly $\alpha$  contrast extraction stage. When this continuum fitting is not performed, the Ly $\alpha$  contrast projection given by Eqn.(7.7-7.9) is not performed either.
- **M**: Modeling of contamination by metal absorption using **quickquasars**.

Depending on whether these systematics are modeled, the notation used to define a mock in this chapter is:

$$SMocks-xcorr-[no]RSD-[no]D-[no]Dm-[no]N-[no]C-[no]CF-[no]M$$

The list of mock realizations used here is given in Tab. 7.1. A total of 32 mocks matching the Stripe 82 geometry were thus created.

These mocks can be broken down into two types. The mocks labelled **Systematics** come all from the same underlying matter density field. For these mocks, Ly $\alpha$  systematics were added progressively to isolate the impact of each of those physical or instrumental effects on the data. Note that the mocks *full-RSD* and *full-noRSD* correspond exactly to the mocks *SMocks-RSD* and *SMocks-noRSD* used in Chap. 6. I also used the matter field of these mocks to calculate the Matter×Void cross-correlation.

I also generated eleven independent mock realizations of the full Stripe 82 dataset, called **Statistics**. This is useful to reduce the statistical uncertainty of the mock prediction itself, and also to validate the error bar calculation done by subsampling. To do so, I created eleven stripes with the same Ly $\alpha$  line-of-sight density and mean separation, and the same footprint as Stripe 82. For these eleven mocks, all Ly $\alpha$  systematics are modeled.

#### Pipeline parameters

The cross-correlation measurements are performed using the pipeline presented in Sec. 7.1.1.

As this study aims to measure the profile of different observables around voids, I limit myself to separations smaller than  $50 h^{-1}\cdot\text{Mpc}$ , a length larger than the maximum radius of the detected voids. The other binning parameters are chosen to provide the best stability for the multipole

Mock type	Suffix	Short name	Realizations
<b>Systematics</b>	noRSD-noD-noDm-noN-noC-noCF-noM	raw-noRSD	1
	noRSD-noD-noDm-noN-C-CF-noM	-	1
	noRSD-noD-noDm-N-C-CF-noM	-	1
	noRSD-noD-noDm-N-C-CF-M	-	1
	noRSD-D-Dm-N-C-CF-M	full-noRSD	1
	RSD-noD-noDm-noN-noC-noCF-noM	raw-RSD	1
	RSD-noD-noDm-noN-C-CF-noM	-	1
	RSD-noD-noDm-N-C-CF-noM	-	1
	RSD-noD-noDm-N-C-CF-M	-	1
	RSD-D-Dm-N-C-CF-M	full-RSD	1
<b>Statistics</b>	noRSD-D-Dm-N-C-CF-M	full11-noRSD	11
	RSD-D-Dm-N-C-CF-M	full11-RSD	11

TABLE 7.1 – List of all Ly $\alpha$  contrast mocks used in this chapter.

derivation: the cross-correlation binning adopted is  $r \in [0, 50] h^{-1} \cdot \text{Mpc}$ ,  $\mu \in [-1, 1]$ ,  $N_r = 25$  and  $N_\mu = 50$ . As detailed in Sec. 7.1.1, the multipole computation is performed applying a Simpson's integration scheme after a linear rebinning to  $N_\mu = 200$  along the  $\mu$  direction. Additionally, the multipole computation is done for separations  $r > 7h^{-1} \cdot \text{Mpc}$ . For the two-dimensional plots, the binning chosen is such that  $r_{\parallel} \in [-50, 50] h^{-1} \cdot \text{Mpc}$ ,  $r_{\perp} \in [0, 50] h^{-1} \cdot \text{Mpc}$ ,  $N_{r_{\parallel}} = 50$ , and  $N_{r_{\perp}} = 25$ .

## 7.2.2 Matter void cross-correlation on mocks

I first measured the Matter $\times$ Void cross-correlation on the density field associated to mocks, to test the cross-correlation pipeline. We want to follow the logic of construction of the matter density field presented in Sec. 6.2.2 and test the impact of each step on the cross-correlation measurement and its multipoles.

The matter density field  $\delta_m$  is computed from the mock's large-scale matter field  $\delta_L$ . In addition,  $\delta_m$  is Gaussian smoothed with  $d_{\text{filter}} = 5 h^{-1} \cdot \text{Mpc}$  to remove small-scale numerical effects. The final matter density field obtained is adapted from Eqn. 3.13:

$$\delta_m = \left[ \exp \left( G(z) \times \delta_L - \frac{\tilde{\sigma}_L^2}{2} \right) - 1 \right] \otimes \text{Gauss}(d_{\text{filter}}). \quad (7.15)$$

To measure the impact of each step in the creation of this field, I realized four declinations with increased complexity. These matter field boxes, called by the prefix *SMocks-xcorr-matter* are the following:

- **noG-noI-noS**: This is a cubic box, with the field  $\delta_m$  as in Eq. 7.15 without applying the factor  $G(z)$ . Therefore, this box has a different geometry from Stripe 82 geometry and corresponds to the matter field at  $z = 0$ .
- **noG-noI-S**: Same as *noG-noI-noS*, but with a geometry similar to Stripe 82.
- **noG-I-S**: Same as *noG-noI-S*, but realising linear interpolation (different from Sec. 6.2.2) when extracting the matter map. Since the  $(\text{RA}, \text{DEC}, z) \rightarrow (X, Y, Z)$  transform used for tomographic map-making does not give the same Cartesian coordinates as the matter density field, this interpolation is required.

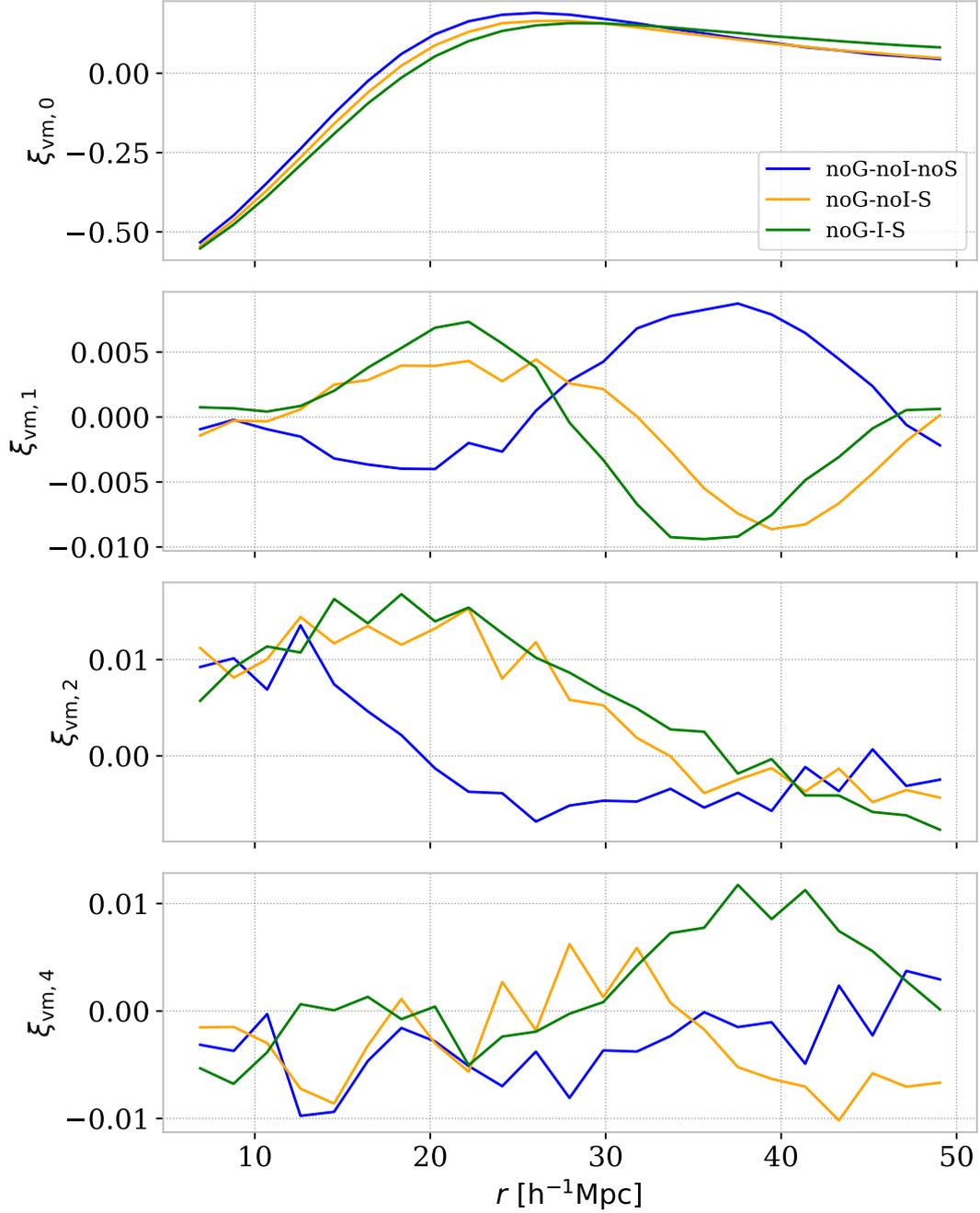


FIGURE 7.5 – Multipoles of the Matter  $\times$  Void cross-correlation for the *SMocks-xcorr-matter -noG-noI-noS*, *-noG-noI-S*, and *-noG-I-S* boxes. The same void finding and cross-correlation parameters are adopted for all these boxes.

- **G-I-S**: This version corresponds to the  $\delta_m$  field as defined by Eqn. 7.15. It is computed at the same redshifts as the mock tomographic map.

The three boxes *SMocks-xcorr-matter -noG-noI-noS*, *-noG-noI-S*, and *-noG-I-S*, all represent the matter density field at  $z = 0$ . For these three fields, I applied the same spherical void finder parameters:  $\delta_{\text{th}} = -0.6$  and  $\delta_{\text{av}} = -0.5$ . For the box *SMocks-xcorr-matter-G-I-S*, the density field is expressed at the redshifts of the Stripe 82 QSOs. The absolute values of this field are smaller. Consequently, I chose the parameters  $\delta_{\text{th}} = -0.3$  and  $\delta_{\text{av}} = -0.2$  in this particular case.

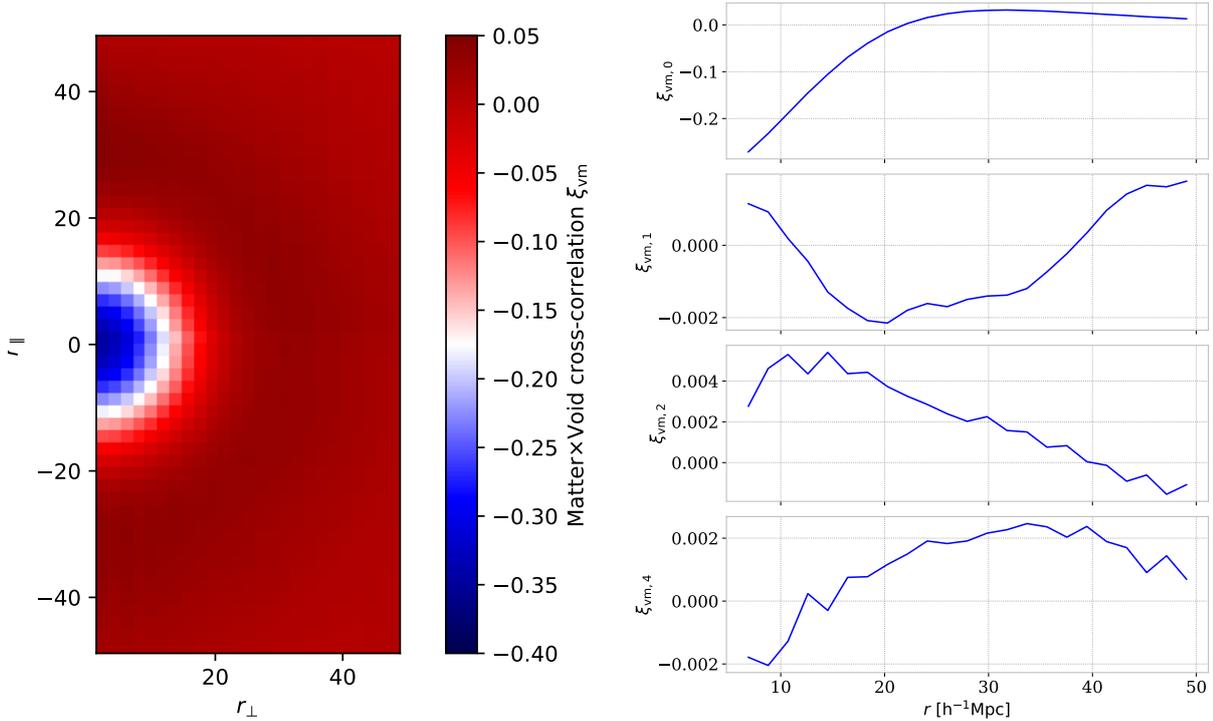


FIGURE 7.6 – (left) 2D representation of the cross-correlation for the *SMocks-xcorr-matter-G-I-S* box. (right) Multipoles of the associated cross-correlation.

For all boxes, the lines-of-sight are drawn from the density field by a uniform and equidistant sampling every 2 pixels. I tested that results are unchanged when extracting lines-of-sight at every pixels.

The measurement of the Matter $\times$ Void cross-correlation and its multipoles is done with the parameters given in Sec. 7.2.1. The multipoles for the three boxes expressed at  $z = 0$  are plotted in Fig. 7.5. The results for the matter field *SMocks-xcorr-matter-G-I-S* are plotted in Fig. 7.6. Since the *SMocks-xcorr-matter-G-I-S* box is at higher redshift and the void finder searches shallower voids, the absolute value of the monopole is lower.

The main conclusion we can draw from these two figures is that all the multipoles  $|\xi_{v\alpha,i}|$  (for  $i = 1, 2, 4$ ) are lower than 4% of  $|\xi_{v\alpha,0}|$ . None of the considered effects generates a spurious multipole with amplitude large enough to significantly bias our measurement. In Fig. 7.5, we can note that the dipole  $\xi_{v\alpha,1}$  tends to deviate from zero at scale  $r \gtrsim 30 h^{-1}\cdot\text{Mpc}$ . We interpret this residual dipole to be due to the box geometry since the its sign is reversed when switching to a Stripe-like geometry. The largest residual effect on multipoles is seen on the quadrupole mostly at small scales (approximately 3% of the monopole for  $r \lesssim 30 h^{-1}\cdot\text{Mpc}$ ) for all mocks with Stripe-like geometries. It is also visible, but to a smaller extent, in the cubic box. Finally, the measured hexadecapoles are compatible with fluctuations around zero which is expected. Even if residual effects are present, we can conclude from these examples that the cross-correlation pipeline can reconstruct the void shapes in the matter density field, with a precision suitable for our purpose.

### 7.2.3 Impact of systematic effects on the Lyman- $\alpha$ $\times$ Void cross-correlation

We now study how some instrumental and astrophysical systematics impact the Ly $\alpha$   $\times$  Void cross-correlation, using the **Systematics** mocks given in Tab. 7.1. For each mock, we compute a tomographic map from which we derive a void catalog, the corresponding Ly $\alpha$   $\times$  Void cross-correlation  $\xi_{v\alpha}$  and its multipole expansion. Fig. 7.7 to 7.10 represent the function  $r \times \xi_{v\alpha}$  in the  $(r_{\parallel}, r_{\perp})$  plane, and the multipoles of  $\xi_{v\alpha}$ . Here are some conclusions one can draw from these plots:

- **Monopole:** The impact of RSD on the monopole  $\xi_{v\alpha,0}$  is small compared to error bars, for all mocks (in average 5% in relative difference).
- **Dipole:** Given the error bars calculated with the sub-sampling, the dipole  $\xi_{v\alpha,1}$  remains compatible with zero for separations  $r > 10 h^{-1} \cdot \text{Mpc}$ . However, there is an indication for an increase of  $\xi_{v\alpha,1}$  at  $r \lesssim 20 h^{-1} \cdot \text{Mpc}$ , after including the effect of continuum fitting and noise.
- **Quadrupole:** The quadrupoles  $\xi_{v\alpha,2}$  in the RSD and noRSD cases can be distinguished for all mocks, given the statistical error bars. This means that the Ly $\alpha$   $\times$  Void cross-correlation contains information about velocity flows, even when including systematics. However, it is important to note that the quadrupole of the mocks without RSD is not null. The numerical tests performed in previous section show that this signal cannot be an artifact of the void finder or the multipole calculation. We interpret this non-null quadrupole as originating from the Ly $\alpha$  tomographic procedure and the intrinsic anisotropies of the Ly $\alpha$  line-of-sight distribution. Unfortunately, this residual quadrupole is concurrent with the RSD effect. It must therefore be taken into account in order to interpret the RSD. I called this the **tomographic effect** and I will come back in detail in Sec. 7.3.2.
- **Hexadecapole:** Finally, the hexadecapoles  $\xi_{v\alpha,4}$  are not null, especially for *raw* mocks. Given the statistical error bars, the presence of RSD does not impact significantly  $\xi_{v\alpha,4}$ .

A more specific comparison of the monopole and quadrupole mocks with RSD is pictured in Fig. 7.11. In this figure, the impact of continuum (both quasar continuum modeling and fitting) is obtained using the *SMocks-xcorr-RSD-noD-noDm-noN-C-CF-noM* mock. I created this mock by using a very high exposure time in *quickquasars* so that the impact of continuum and noise are decoupled. The void finder does not work properly with this mock (void centers are significantly offset): this is probably because the Ly $\alpha$  tomography map gives too important weights to some Ly $\alpha$  contrast. Consequently, I used voids detected on the *raw-RSD* mock and correlated them with the Ly $\alpha$  contrast of this *SMocks-xcorr-RSD-noD-noDm-noN-C-CF-noM* mock. This specific measure is called from now on *raw-RSD-continuum*.

In Fig. 7.11, the continuum mainly impacts the monopole. At small  $r$ , the monopole slightly decreases: this signal loss is introduced by the continuum modeling and its fitting. For larger measured separations  $r$ , the continuum effect increases the absolute value of the cross-correlation: approximately a 35% difference at  $r \sim 30 h^{-1} \cdot \text{Mpc}$ . This effect is well known in the calculation of the Ly $\alpha$   $\times$  Ly $\alpha$  and Ly $\alpha$   $\times$  QSO correlations.

To make sure we understand this effect, I carried out an explicit comparison with the case of the Ly $\alpha$   $\times$  QSO cross-correlation in [15]. I calculated Ly $\alpha$   $\times$  QSO on the *SMocks-xcorr-RSD-noD-noDm-noN-C-CF-noM* mock, to obtain the same separation binning and statistics as the Ly $\alpha$   $\times$  Void cross-correlation. It is known from recent BAO studies [14, 15, 16] that the bins mostly

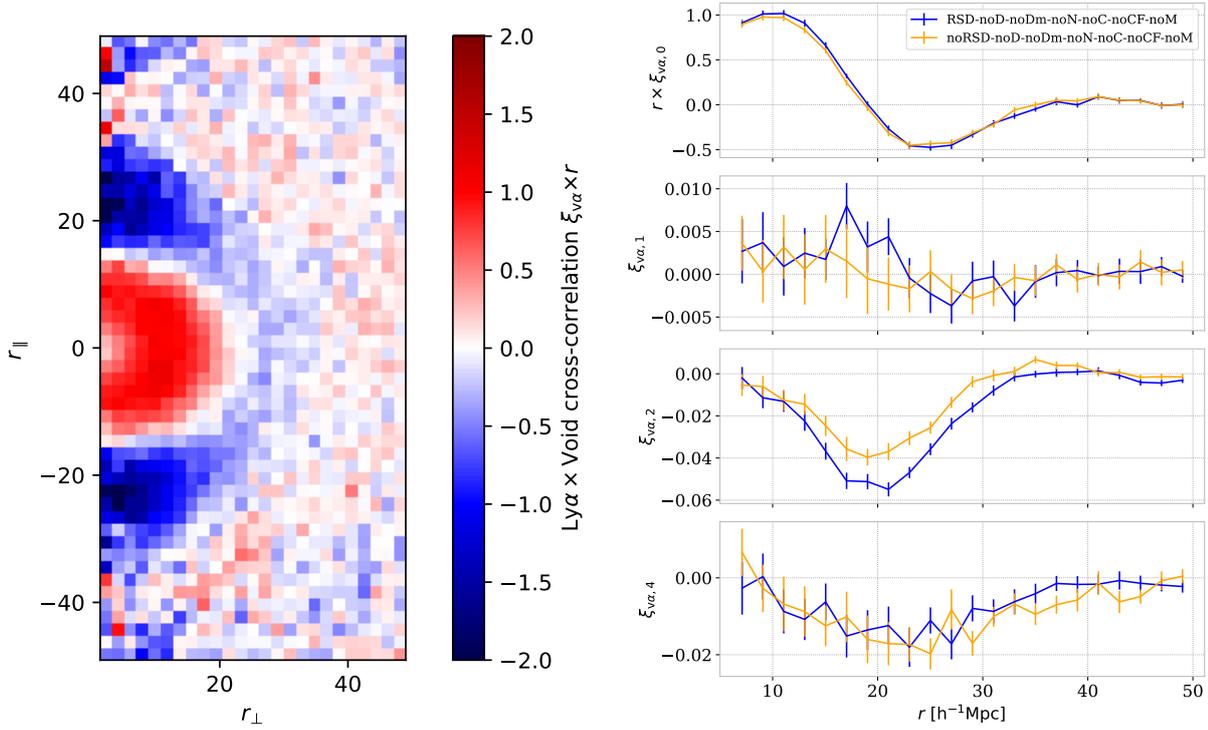


FIGURE 7.7 – (left) 2D representation of the cross-correlation for *SMocks-xcorr-RSD-noD-noDm-noN-noC-noCF-noM* mock (i.e. *raw-RSD* mock). (right) Multipoles of the associated cross-correlation (blue) and comparison with the mock realization without RSD modeling (orange).

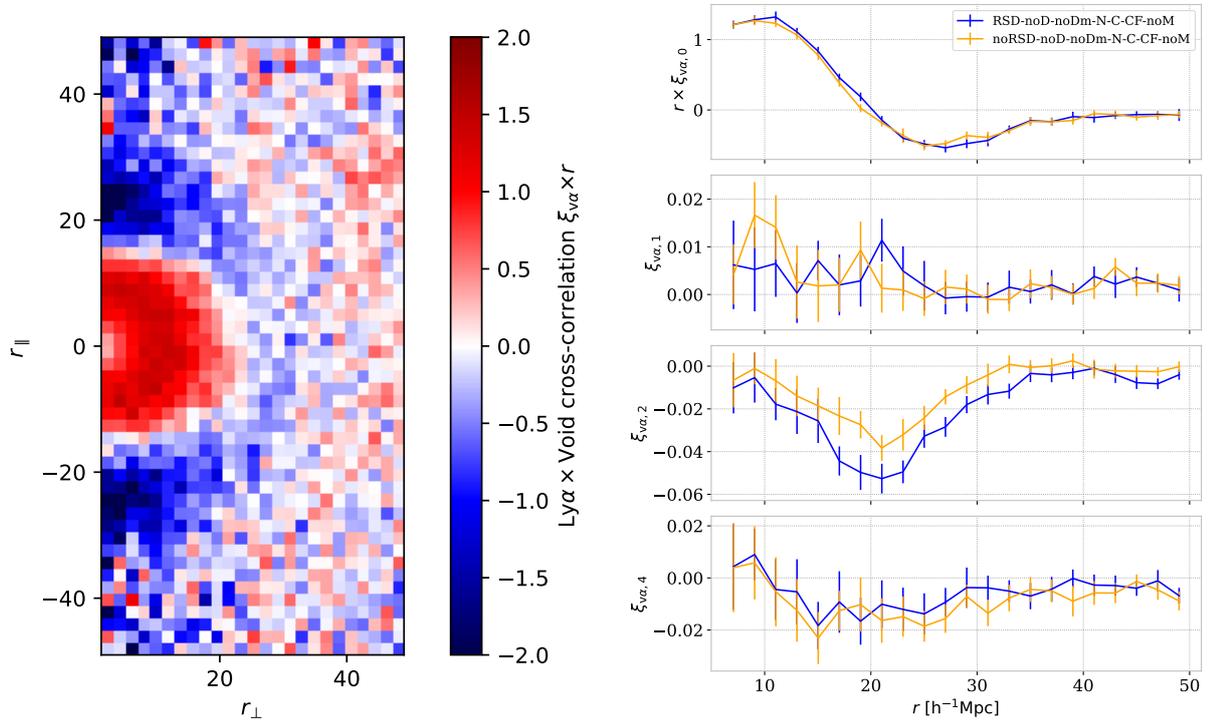


FIGURE 7.8 – (left) Same as Fig. 7.7 for the *SMocks-xcorr-RSD-noD-noDm-N-C-CF-noM* mock (i.e. adding the effect of noise and quasar continuum fitting).

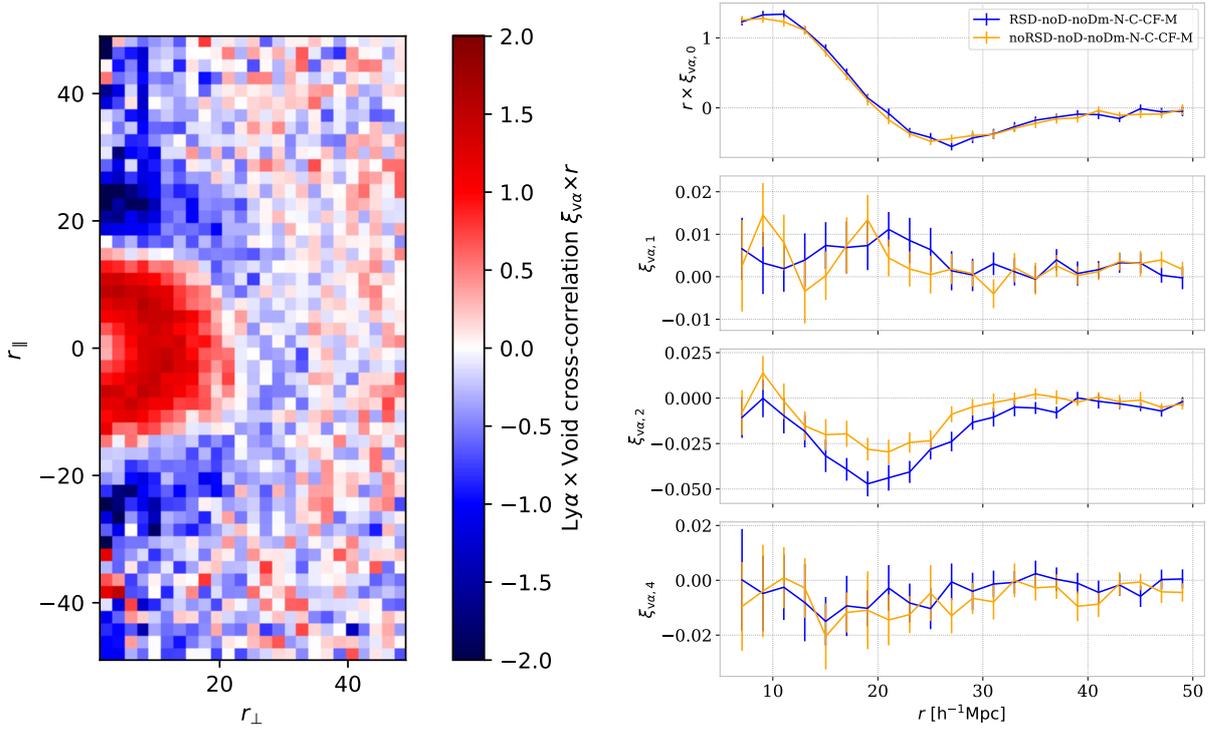


FIGURE 7.9 – Same as Fig. 7.8 for the *SMocks-xcorr-RSD-noD-noDm-N-C-CF-M* mock (i.e. adding the effect of metals).

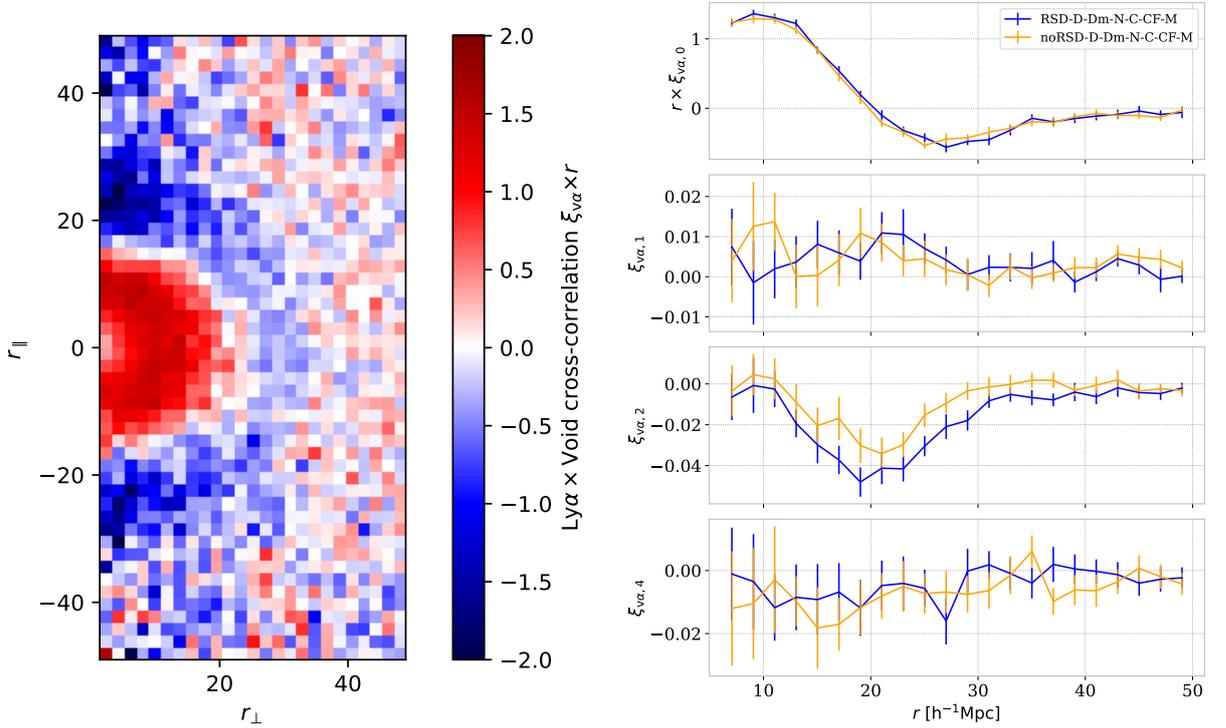


FIGURE 7.10 – Same as Fig. 7.9 for the *SMocks-xcorr-RSD-D-Dm-N-C-CF-M* mock (i.e. adding DLAs and their masking).

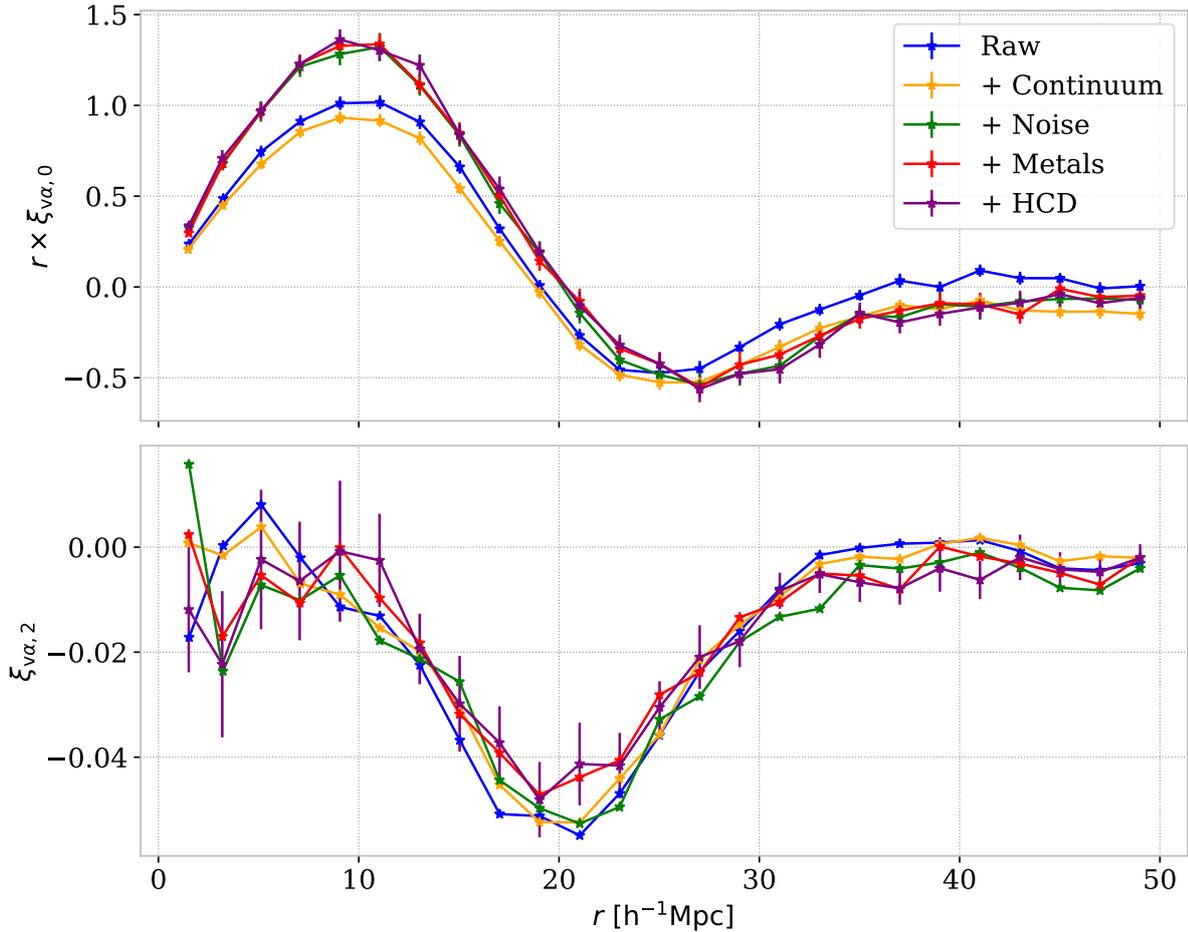


FIGURE 7.11 – Measured monopoles and quadrupoles from the **Systematics** mocks, including different instrumental and astrophysical effects. Statistical error bars are estimated using subsampling variance (for the quadrupole they are shown only for the purple curve).

affected by the continuum fitting are the small transverse separations  $r_{\perp}$ . Fig. 7.12 shows the comparison between the mock *raw-RSD* and the *raw-RSD-continuum* cross-correlations for QSO and void, averaged in a range  $3 < r_{\perp} < 11 h^{-1}\cdot\text{Mpc}$ .

The  $\text{Ly}\alpha \times \text{QSO}$  cross-correlation measurement is in agreement with the one shown in [15] with larger statistical uncertainties. On the  $r_{\perp}$  range chosen, it appears that the  $\text{Ly}\alpha \times \text{QSO}$  cross-correlation is more impacted by the continuum fitting than the  $\text{Ly}\alpha \times \text{Void}$ . Moreover, this impact is more asymmetric between  $r_{\parallel} < -20 h^{-1}\cdot\text{Mpc}$  and  $r_{\parallel} > 20 h^{-1}\cdot\text{Mpc}$  for QSOs than for voids. In conclusion, it seems that the continuum impact, although not negligible, is less strong in the case of the  $\text{Ly}\alpha \times \text{Void}$  cross-correlation.

Going back to Fig. 7.11, when noise is added, additional small-scale fluctuations in  $\delta_F$  significantly change the associated void catalog, increasing the number of small, spurious voids. This results in an increase of  $|\xi_{\text{v}\alpha,0}|$  at small separations. Furthermore, the noise changes the weights of low redshift pixels relative to high redshift pixels, so the effective bias of the correlation is larger for mocks with noise.

Fig. 7.11 shows that metals only marginally impact the correlation function. HCD systems do not modify the measured cross-correlation, given the error bars. This is expected since identified voids are located away from HCD objects.

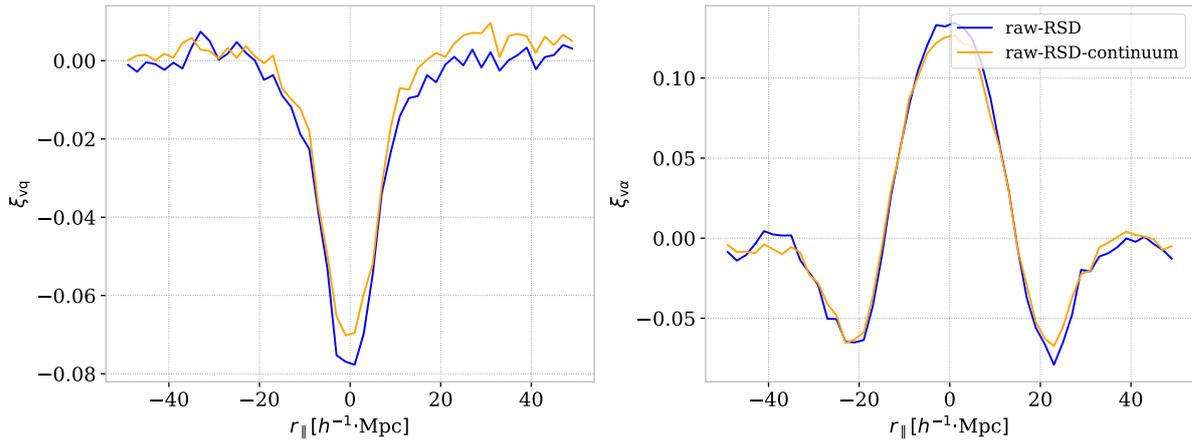


FIGURE 7.12 – (left) Ly $\alpha$ ×QSO cross-correlation. (right) Ly $\alpha$ ×Void cross-correlation. Both cross-correlations are represented with  $A = (r_{\parallel}, r_{\perp})$  binning and averaged in the  $3 < r_{\perp} < 11 h^{-1} \cdot \text{Mpc}$  range for the *raw-RSD* mock and the *raw-RSD-continuum* measurement.

As a conclusion, for the data set considered in this study, all instrumental and astrophysical effects on the quadrupole are relatively minor with respect to statistical fluctuations. This result is very encouraging to use the quadrupole of the Ly $\alpha$ ×Void cross-correlation as a probe of RSD.

### Dependence on the void finder parameters

I performed tests on *raw-noRSD* mock variations to characterize the impact of the Ly $\alpha$  tomography and void finder parameters. The effect of these parameters on the Ly $\alpha$ ×Void cross-correlation multipoles is represented in Fig. 7.13.

First, I change simultaneously the transverse and longitudinal correlation lengths  $L_{\perp}$  and  $L_{\parallel}$ , used to compute tomographic maps, from  $10$  to  $16 h^{-1} \cdot \text{Mpc}$ . As shown on the left of Fig. 7.13, the multipoles are strongly influenced by the correlation lengths. This is expected, as  $L_{\perp}$  and  $L_{\parallel}$  ultimately dictate the size of identified voids. This impact can be reduced by rescaling the cross-correlation separation  $r$  with the correlation lengths.

The impact of the void finder parameter  $\delta_{\text{av}}$  is given in Fig. 7.13 (right). Keeping the threshold of the spherical void finder constant, I modified  $\delta_{\text{av}}$  from  $0.12$  to  $0.16$ . Compared to the correlation lengths, the impact of this parameter is smaller. As the size of the selected voids is increased for  $\delta_{\text{av}} = 0.16$ , the voids will be on average deeper, which results in an increase of the absolute value of the monopole at small separation.

These two tests were performed as well on the *raw-RSD* mock: I checked that the quadrupole difference due to the RSD remains discernible for each parameter variation.

Finally, I conducted an exploratory test which consisted in varying the value of  $\Omega_{\text{m},0}$  in the distances and cross-correlation computations. The purpose of this test was to potentially see if the Ly $\alpha$ ×Void cross-correlation was sensitive to this parameter. Varying this parameter enlarges the void size, but it is less significant than the correlation lengths. Therefore, this test is not conclusive as long as the impact of the correlation lengths is not perfectly controlled.

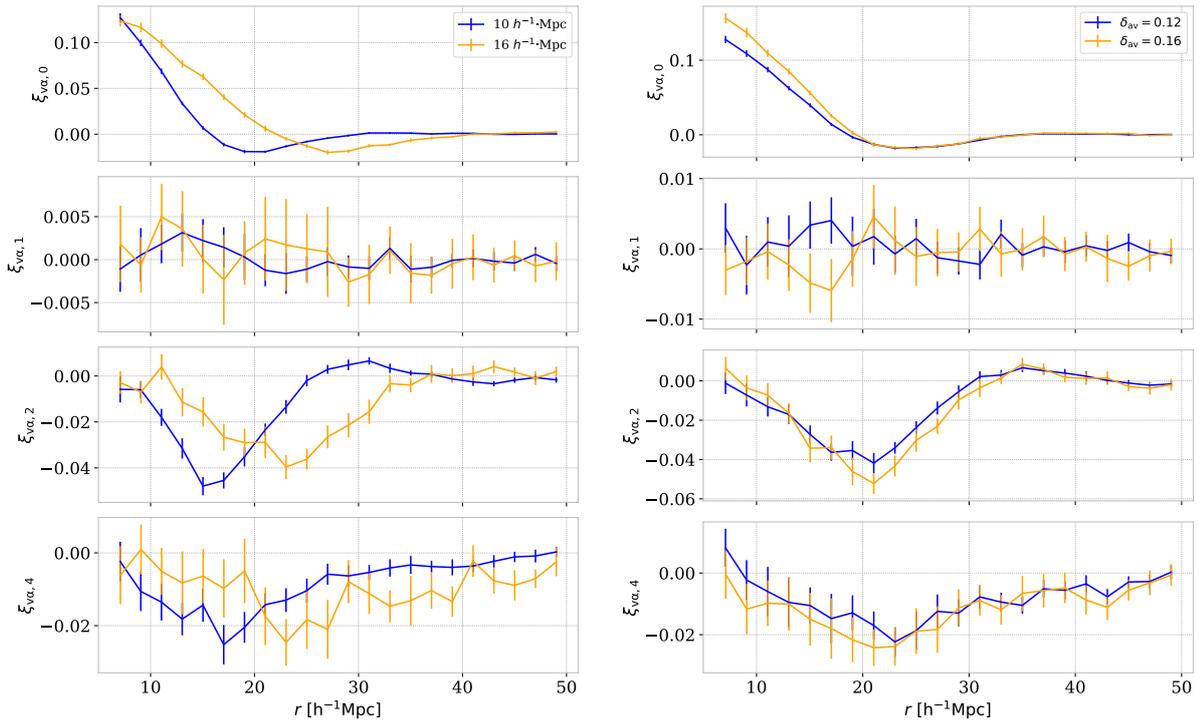


FIGURE 7.13 – Impact of analysis parameters on the *raw-noRSD* mock multipoles. (left) Variation of both  $L_{\perp}$  and  $L_{\parallel}$  correlation lengths from 10 to 16  $h^{-1}$ ·Mpc. (right) Variation of spherical void finder parameter  $\delta_{\text{av}}$  from 0.12 to 0.16.

## 7.3 Modeling the Lyman- $\alpha$ $\times$ Void cross-correlation

### 7.3.1 Redshift space distortions in Lyman- $\alpha$ $\times$ Void

I presented the effect of RSD on the power spectrum of a Boolean tracer in Sec. 1.3.3.1, and in Sec. 2.2.2, for the case of the Ly $\alpha$  forest. In the latter case, the amplitude of the RSD effect is proportional to the Ly $\alpha$  velocity bias  $b_\eta$ . In this section, the impact of RSD is studied in the framework of voids. I propose a very simplified, but effective model to describe the velocity behavior of the IGM gas around voids. In particular, I rely on several hypotheses. The development of this model is based on the articles [4, 5, 21].

#### Coordinates

To develop this model, I use the Eulerian approach detailed in Sec. 1.3.3.1. I define the physical velocity field  $\vec{u}$ , matter density contrast  $\delta$ , both expressed as a function of comoving position  $\vec{x}$ . In real space, the comoving position of a void center is noted  $\vec{X}$  and its physical velocity is expressed by  $\vec{U}$ . I can then define  $\vec{r} = \vec{x} - \vec{X}$  the comoving void-gas separation in real space.

In redshift space, I define the comoving void position  $\tilde{X}$ , the comoving gas position  $\tilde{x}$ , and the comoving void-gas separation  $\tilde{s} = \tilde{x} - \tilde{X}$ .

In this section, I consider that the position of the voids is known precisely, i.e. that the void centers are perfectly reconstructed.

I define the void direction in real space  $\hat{X} = \frac{\vec{X}}{|\vec{X}|}$ . I consider the cylindrical coordinate basis  $(\hat{X}, \hat{Y}, \phi)$  to decompose transverse and parallel directions in real space. In this basis, I can define  $r_\parallel = \vec{r} \cdot \hat{X}$ ,  $r_\perp = \vec{r} \cdot \hat{Y}$ ,  $r = \sqrt{r_\parallel^2 + r_\perp^2}$ , and  $\mu = \cos(\theta) = r_\parallel/r$ . The same projections are defined for the comoving void-gas separation in redshift space.

#### Velocity profile

*Hyp. 1:* I suppose that the void shape is spherical, and that around the void, in real space, the physical velocity field  $\vec{u}$  follows linear perturbation theory.

This simplifying hypothesis is the most important in our model. Individual voids are clearly not spherical. However, the spherical shape assumption is relevant since, in the end, we measure averages over a large number of voids: the average void profile is spherical. The linear assumption for  $\vec{u}$  is qualitatively justified because non-linearities are less intense around voids than around halos, and we consider here high redshifts ( $z > 2$ ).

In this linear framework, it is possible to use the spherical collapse model to describe the density field around voids. Eqn. 1.132 yields the following linear velocity field:

$$\vec{u} = -\frac{1}{3} \frac{fH}{1+z} \bar{\delta}(r) \vec{r}, \quad (7.16)$$

with the average isotropic matter density contrast  $\bar{\delta}(r)$  given by

$$\bar{\delta}(r) = \frac{3}{r^3} \int_0^r \delta(r') r'^2 dr'. \quad (7.17)$$

#### RSD transformation

It is then possible to relate the real space and the redshift space by starting from Eqn. 1.135 which gives:

$$\vec{s} = \vec{r} + (1+z) \frac{\hat{X} \cdot (\vec{u} - \vec{U})}{H(z)} \hat{X}. \quad (7.18)$$

Here, we use the distant-observer approximation, i.e.  $\vec{x}$  and  $\vec{X}$  are considered parallel since  $|r| \ll |X|$ .

*Hyp. 2:* The void center velocity is negligible compared to the IGM gas [22].

Due to hypothesis 2, the void speed can be neglected compared to the gas elements:

$$\vec{s} = \vec{r} + (1+z) \frac{\hat{X} \cdot \vec{u}}{H(z)} \hat{X}. \quad (7.19)$$

*Hyp. 3:* The redshift space to real space mapping locally conserves the optical depth [21] (similarly to Sec. 2.2.2):

$$\tau_\alpha(\vec{r}) d^3 \vec{r} = \tau_\alpha^s(\vec{r}) d^3 \vec{s} = \tau_\alpha^s(\vec{r}) \left| \frac{d^3 \vec{s}}{d^3 \vec{r}} \right| d^3 \vec{r}, \quad (7.20)$$

where the <sup>s</sup> exponent indicates that the function is expressed in redshift space. Merging equations 7.19 and 7.16, the transformation is simplified as follow:

$$\begin{aligned} \vec{s} &= \vec{r} - \frac{1}{3} (\hat{X} \cdot \vec{r}) f \bar{\delta}(r) \hat{X} \\ &= \vec{r} - \frac{1}{3} r_{\parallel} f \bar{\delta}(r) \hat{X}. \end{aligned} \quad (7.21)$$

The Jacobian of this transformation is given by

$$\begin{aligned} \left| \frac{d^3 \vec{s}}{d^3 \vec{r}} \right| &= \left| \begin{array}{cc} \frac{ds_{\parallel}}{dr_{\parallel}} & \frac{ds_{\perp}}{dr_{\perp}} \\ \frac{ds_{\parallel}}{dr_{\perp}} & \frac{ds_{\perp}}{dr_{\parallel}} \end{array} \right| \\ &= \left| \begin{array}{cc} 1 - \frac{f}{3} \bar{\delta}(r) - \frac{f}{3} r_{\parallel} \frac{\partial \bar{\delta}(r)}{\partial r_{\parallel}} & 0 \\ -\frac{f}{3} r_{\parallel} \frac{\partial \bar{\delta}(r)}{\partial r_{\perp}} & 1 \end{array} \right| \\ &= 1 - \frac{f}{3} \bar{\delta}(r) - \frac{f}{3} r_{\parallel} \frac{\partial r}{\partial r_{\parallel}} \frac{d\bar{\delta}(r)}{dr} \\ &= 1 - \frac{f}{3} \bar{\delta}(r) - \frac{f}{3} \mu^2 r \frac{d\bar{\delta}(r)}{dr}. \end{aligned} \quad (7.22)$$

The derivative of the average isotropic matter density contrast  $\bar{\delta}(r)$  with respect to  $r$  is obtained using Eqn. 7.17:

$$\frac{d\bar{\delta}(r)}{dr} = \frac{3}{r} (\delta(r) - \bar{\delta}(r)). \quad (7.23)$$

Merging Eqn. 7.22 and Eqn. 7.23, the Jacobian simplifies as follow:

$$\left| \frac{d^3 \vec{s}}{d^3 \vec{r}} \right| = 1 - \frac{f}{3} \bar{\delta}(r) - f \mu^2 (\delta(r) - \bar{\delta}(r)). \quad (7.24)$$

Using equation 7.20, I can express  $\tau_\alpha^s(\vec{r})$  in function of  $\tau_\alpha(r)$ :

$$\begin{aligned} \tau_\alpha^s(\vec{r}) &= \tau_\alpha(r) \left| \frac{d^3 \vec{r}}{d^3 \vec{s}} \right| \\ &= \tau_\alpha(r) \left[ 1 + \frac{f}{3b} \bar{\delta}(r) + \frac{f}{b} \mu^2 (\delta(r) - \bar{\delta}(r)) \right]. \end{aligned} \quad (7.25)$$

### Link with the Ly $\alpha$ contrast

I follow the same derivation as Sec. 2.2.2 and in [21], i.e. considering only long-wavelength modes of the Ly $\alpha$  contrast, and expanding the fraction of transmitted flux  $F$  to first order as a function of  $\tau$ . I obtain the relation between real and redshift space fraction of transmitted flux (analog of Eqn. 2.37):

$$F[\tau_\alpha^s(\vec{r})] = F[\tau_\alpha(r)] + b_\eta f \bar{F} \left[ \frac{1}{3} \bar{\delta}(r) + \mu^2 (\delta(r) - \bar{\delta}(r)) \right]. \quad (7.26)$$

In term of Ly $\alpha$  contrast  $\delta_F$ , this relation writes:

$$\delta_F^s(\vec{r}) = \delta_F(r) + b_\eta f \left[ \frac{\bar{\delta}(r)}{3} + \mu^2 (\delta(r) - \bar{\delta}(r)) \right]. \quad (7.27)$$

### Ly $\alpha$ $\times$ Void cross-correlation

The Ly $\alpha$   $\times$  Void cross-correlation is defined by Eqn. 7.1. In the absence of RSD,  $\xi_{v\alpha}(r)$  is isotropic. I define the average Ly $\alpha$   $\times$  Void cross-correlation  $\bar{\xi}_{v\alpha}$  by

$$\bar{\xi}_{v\alpha}(r) = \frac{3}{r^3} \int_0^r \xi_{v\alpha}(\vec{r}') r'^2 d^3 r'. \quad (7.28)$$

*Hyp. 4*: The real-space (isotropic) Ly $\alpha$   $\times$  Void cross-correlation linearly follows the matter density contrast as suggested in [23].

This hypothesis is the most important of this derivation. It was verified in [23] on N-body simulation at scales near to  $50 h^{-1}$ -Mpc and redshift  $z = 0.17$  for the Galaxy  $\times$  Void cross-correlation.

In our case, we consider smaller distances on the one hand, but higher redshifts on the other hand. A full justification of *Hyp. 4* in the case of Ly $\alpha$  would require using hydrodynamic simulations. For now, this assumption is assumed to be valid, and consequently, I assume the following relationship:

$$\xi_{v\alpha}(r) = b\delta(r). \quad (7.29)$$

The **void-Ly $\alpha$  bias**  $b$  is defined by this relation, or equivalently by

$$b = \left. \frac{\partial \xi_{v\alpha}(r)}{\partial \delta(r)} \right|_{\xi_{v\alpha}=0}. \quad (7.30)$$

From the definition of average quantities, it also implies that  $\bar{\delta}$  and  $\bar{\xi}_{v\alpha}$  are connected by the same relation. Taking into account this additional bias and applying the definition of Ly $\alpha$   $\times$  Void cross-correlation, Eqn. 7.27 yields:

$$\xi_{v\alpha}^s(\vec{r}) = \xi_{v\alpha}(r) + \beta \left[ \frac{\bar{\xi}_{v\alpha}(r)}{3} + \mu^2 (\xi_{v\alpha}(r) - \bar{\xi}_{v\alpha}(r)) \right], \quad (7.31)$$

where I have defined the **void-Ly $\alpha$  RSD parameter**  $\beta$  by

$$\beta = \frac{b_\eta f}{b}. \quad (7.32)$$

Within this model, it is possible to link this RSD parameter  $\beta$  to the RSD parameter defined by Eqn. 2.39 such that  $\beta_\alpha = (b/b_\alpha)\beta$ .

### Multipole expansion

The multipole expansion is defined by Eqn. 7.14. From Eqn. 7.31, only the quadrupole and the monopole are non-vanishing:

$$\begin{aligned}\xi_{\text{v}\alpha,0}^s(r) &= \left(1 + \frac{\beta}{3}\right) \xi_{\text{v}\alpha}(r), \\ \xi_{\text{v}\alpha,2}^s(r) &= \left(\frac{2\beta}{3}\right) (\xi_{\text{v}\alpha}(r) - \bar{\xi}_{\text{v}\alpha}(r)).\end{aligned}\tag{7.33}$$

The radial dependence of  $\xi_{\text{v}\alpha,2}^s$  is therefore entirely governed by the monopole, i.e. the void profile:

$$\xi_{\text{v}\alpha,2}^s(r) = \left(\frac{2\beta}{3 + \beta}\right) (\xi_{\text{v}\alpha,0}^s(r) - \bar{\xi}_{\text{v}\alpha,0}^s(r)).\tag{7.34}$$

This relation is formally identical to the one obtained in simple models of the Galaxy  $\times$  Void cross-correlation [5] where the velocity flow generates a quadrupole. In my case, the velocity bias  $b_\eta$  impedes a straightforward inference of the logarithmic growth rate of linear perturbations  $f$ .

## 7.3.2 Tomographic effect

As seen in Sec. 7.2.3, the measured Ly $\alpha$   $\times$  Void cross-correlation has a non-zero quadrupole even in the absence of RSD. This feature comes from applying the Wiener filter to the anisotropic geometry of the lines-of-sight, which affects the reconstructed void positions. Indeed, the average flux contrast of the tomographic map built with the Wiener filter is smaller at locations further away from lines-of-sight. It decreases the efficiency of the void finder at these parts of the map and displaces the average reconstructed positions of void centers. As a result, the voids appear closer towards the nearest line-of-sight with respect to their correct positions. This effect, hereafter labeled as **tomographic effect**, has a geometrical origin; hence they are best taken into account in mock realizations. This section aims to model the tomographic effect and perform tests on the mocks detailed in Sec. 7.2.

More precisely, the tomographic effect can emerge from two potential mechanisms:

- **Void position:** The reconstructed center of identified voids is offset systematically towards the lines-of-sight.
- **Finder efficiency:** True voids are missed away from lines-of-sight, which means that the void finder efficiency decreases as a function of  $r_\perp$ .

Here, I present two toy-models emphasizing each mechanism to describe the tomographic effect semi-analytically. In the end, both mechanisms have a similar effect: the distribution of voids with respect to lines-of-sight is shifted towards smaller  $r_\perp$ .

### Model 1: void position

Similarly to the case of RSD, I refer to  $\vec{r}$  as the space of real separations, and  $\vec{s}$  as the space of observed separations. In the case of RSD,  $\vec{s}$  is different from  $\vec{r}$  because RSD affects

$r_{\parallel}$ . On the other hand, on the case of the tomographic effect, the difference between  $\vec{s}$  and  $\vec{r}$  is due to a change in  $r_{\perp}$ . In this model, I assume that the tomographic effect modifies the observed transverse separation according to the function  $s_{\perp} = g(r_{\perp})$ . I define the  $g$  function such that  $r_{\perp} = g^{-1}(s_{\perp}) \equiv s_{\perp} + \epsilon(s_{\perp})$ . I assume that the correction is small:  $\epsilon(s_{\perp}) \ll s_{\perp}$ . This approximation is justified in Fig. 7.14 using mocks. Since reconstructed voids are, on average, closer to lines-of-sight than true voids,  $\epsilon$  is positive.

For simplification, I first develop this model without considering RSD ( $b_{\eta} = 0$ ). The transformation between  $\vec{s}$  and  $\vec{r}$  is then determined only by the tomographic effect:

$$\vec{s} = s_{\parallel} \hat{X} + s_{\perp} \hat{Y} = r_{\parallel} \hat{X} + (r_{\perp} - \epsilon(r_{\perp})) \hat{Y}. \quad (7.35)$$

Then, the density contrast as a function of separation  $\vec{s}$  can be expressed as

$$\delta(\vec{s}) = \delta \left( \sqrt{r_{\parallel}^2 + (r_{\perp} - \epsilon(r_{\perp}))^2} \right). \quad (7.36)$$

The approximation  $\epsilon(r_{\perp}) \ll s_{\perp}$  for all separations yields:

$$\sqrt{r_{\parallel}^2 + (r_{\perp} - \epsilon(r_{\perp}))^2} \simeq r - \epsilon(r_{\perp}) \frac{r_{\perp}}{r}. \quad (7.37)$$

I perform a Taylor-expansion for  $\delta$ , to first order in  $\epsilon$  such that:

$$\delta(\vec{s}) \simeq \delta(r) - \epsilon(r_{\perp}) \frac{d\delta}{dr} \frac{r_{\perp}}{r}. \quad (7.38)$$

Finally, using the fact that  $r_{\perp} = r\sqrt{1 - \mu^2}$ , and the linearity hypothesis between Ly $\alpha$   $\times$  Void cross-correlation and density field, I can express the Ly $\alpha$   $\times$  Void cross-correlation after this change of coordinate:

$$\frac{\xi_{v\alpha}^s(\vec{r})}{b} \simeq \delta(r) - \epsilon(r_{\perp}) \frac{d\delta}{dr} \sqrt{1 - \mu^2}. \quad (7.39)$$

The order of magnitude of the tomographic effect correction is  $\sim \epsilon/r$ . The  $\mu$ -dependence of the correction is in  $\sqrt{1 - \mu^2}$ , introducing multipoles to all even orders.

Now, including also the RSD effect, the transformation between redshift and real space is given by

$$\vec{s} = \vec{r} - \frac{1}{3} r_{\parallel} f \bar{\delta}(r) \hat{X} - \epsilon(r_{\perp}) \hat{Y}. \quad (7.40)$$

From this Eqn. 7.40, one can compute the Jacobian similarly to Eqn. 7.22. The complete expression for the cross-correlation, assuming linear RSD and  $\epsilon(s_{\perp}) \ll s_{\perp}$ , can be derived as

$$\begin{aligned} \xi_{v\alpha}^s &= \xi_{v\alpha} \\ &+ \beta \left[ \frac{\bar{\xi}_{v\alpha}}{3} + \mu^2 (\xi_{v\alpha} - \bar{\xi}_{v\alpha}) \right] \\ &- \left( \epsilon \sqrt{1 - \mu^2} \right) \frac{d\xi_{v\alpha}}{dr} \\ &- \beta \left( \epsilon \sqrt{1 - \mu^2} \right) \left[ \frac{1}{3} \frac{d\bar{\xi}_{v\alpha}}{dr} + \mu^2 \frac{d(\xi_{v\alpha} - \bar{\xi}_{v\alpha})}{dr} - \frac{2}{r} \mu^2 (\xi_{v\alpha} - \bar{\xi}_{v\alpha}) \right]. \end{aligned} \quad (7.41)$$

The first line is the raw void profile, the second line is the linear RSD correction, the third line is the tomographic effect, and the last line is the first-order approximation to the coupling of both RSD and tomographic effects. The order of magnitude of these coupling terms relative to the void profile is equal to  $\beta\epsilon/r$ , so that it is reasonable to neglect these terms with respect to the others. Under this assumption, even after RSD is included, the tomographic effect is an additive term noted

$$\xi_{\text{tomo}}(\vec{r}) = \epsilon(r_{\perp})\sqrt{1 - \mu^2}\frac{d\xi_{\text{v}\alpha}(r)}{dr}. \quad (7.42)$$

The multipoles  $\xi_{\text{tomo},\ell}$  of this correction are calculated using Eqn. 7.14. I use the change of variable  $\mu^2 = 1 - x^2$ , given that  $d\mu = -xdx/\sqrt{1 - x^2}$  and  $r_{\perp} = rx$ . I define the weighted averages of  $\epsilon$  by

$$\langle\epsilon\rangle_i = \frac{\int_0^1 \epsilon(rx)x^i/\sqrt{1 - x^2}}{\int_0^1 x^i/\sqrt{1 - x^2}}. \quad (7.43)$$

I use the known integrals  $\int_0^1 x^i dx/\sqrt{1 - x^2} = \pi/4, 3\pi/16, 5\pi/32$  for  $i = 2, 4, 6$  respectively. The calculations yield:

$$\begin{aligned} \xi_{\text{tomo},0} &= \frac{\pi}{4} \frac{d\xi_{\text{v}\alpha}}{dr} \langle\epsilon\rangle_2, \\ \xi_{\text{tomo},2} &= \frac{5\pi}{4} \frac{d\xi_{\text{v}\alpha}}{dr} \left( \langle\epsilon\rangle_2 - \frac{9}{8} \langle\epsilon\rangle_4 \right), \\ \xi_{\text{tomo},4} &= \frac{9\pi}{4} \frac{d\xi_{\text{v}\alpha}}{dr} \left( \langle\epsilon\rangle_2 - \frac{15}{4} \langle\epsilon\rangle_4 + \frac{175}{64} \langle\epsilon\rangle_6 \right). \end{aligned} \quad (7.44)$$

We can use mocks to estimate  $\epsilon$ . To achieve this, I create random void catalogs on the *full-RSD* mock. I keep the  $Z$  coordinate of all the voids to preserve the redshift distribution, and create random coordinates in  $X$  and  $Y$  directions. I compute the cumulative distribution function of the transverse void separations to the lines-of-sight for the random voids,  $N_{\text{random}}(r_{\perp})$ , as well as for the observed voids, identified from the tomographic map,  $N(r_{\perp})$ . In the absence of tomographic effects, both cumulative distribution functions would be identical. Thus, the average transverse void displacement  $\epsilon$  can be found from  $N(r_{\perp} + \epsilon(r_{\perp})) = N_{\text{random}}(r_{\perp})$ . The  $\epsilon$  function estimated with this method is represented in black in Fig. 7.14. For  $r \sim 20 - 30 h^{-1}\text{Mpc}$ ,  $\epsilon$  is very slowly varying, with  $\epsilon \sim 2 h^{-1}\text{Mpc}$ . This validates the hypothesis  $\epsilon(s_{\perp}) \ll s_{\perp}$ .

Fig. 7.14 also shows the calculated averages  $\langle\epsilon\rangle_i$  ( $i = 2, 4, 6$ ). They are almost identical. Consequently, according to this toy model, the tomographic effect induces non-zero quadrupoles and hexadecapoles with a common radial dependence which scales essentially like  $d\xi_{\text{v}\alpha}/dr$ . On the other hand, the RSD quadrupole scales like  $(\xi_{\text{v}\alpha} - \bar{\xi}_{\text{v}\alpha})(r)$ . The function  $\bar{\xi}_{\text{v}\alpha}$  is a weighted average of  $\xi_{\text{v}\alpha}(y)$  over  $y \in [0, r]$ , so that there exists an  $h \in [0, r]$  such that  $\bar{\xi}_{\text{v}\alpha}(r) = \xi_{\text{v}\alpha}(r - h)$ . Large values of  $y$  are favored in the weighting scheme (in  $y^2$ ), so that  $h$  is small, and consequently  $\xi_{\text{v}\alpha}(r) - \bar{\xi}_{\text{v}\alpha}(r) = \xi_{\text{v}\alpha}(r) - \xi_{\text{v}\alpha}(r - h)$  is close to  $d\xi_{\text{v}\alpha}/dr$ . Therefore, this toy model qualitatively explains why the shape of the tomographic quadrupole is unfortunately similar to that of the RSD quadrupole.

However, this first toy model has limitations. Mocks show that the shapes of  $\xi_{\text{tomo},2}$  and  $\xi_{\text{tomo},4}$ , although roughly similar, are not identical. Their measured ratio  $|\xi_{\text{tomo},4}/\xi_{\text{tomo},2}|$  is also larger by a factor of two approximately, with respect to the prediction of 7.44. These findings

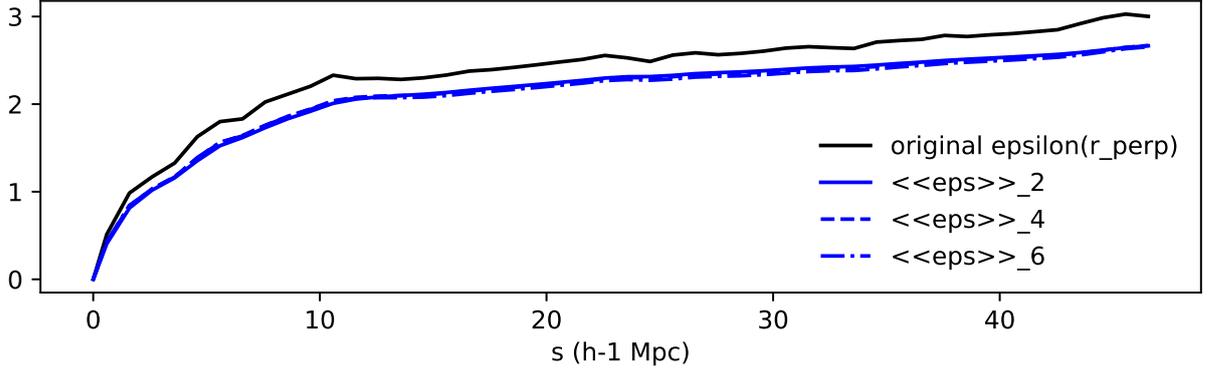


FIGURE 7.14 – Measurement of the  $\epsilon$  function using a randomized void catalog keeping the same redshift distribution. The weighted averages  $\langle \epsilon \rangle_i$  ( $i = 2, 4, 6$ ) are represented in different blue lines.

indicate that the values of multipole ratios obtained by this model are not trustworthy and that we should let them as a free parameter in the final model.

### Model 2: void finder efficiency

Now, I try to emphasize the role of the second effect, related to the void finder efficiency. For simplicity, I consider here only the tomographic effect, independently of the RSD. The loss of efficiency of the void finder away from the lines-of-sight means that the number density of voids is a function of distance to the nearby lines-of-sight; or equivalently, around a given void, the number density of lines-of-sight,  $n_{\text{los}}(r_{\perp})$ , is a decreasing function of  $r_{\perp}$ . The measured cross-correlation is then given by

$$\xi_{\text{v}\alpha}^s(\vec{r}) = \xi_{\text{v}\alpha}(r) \times \frac{n_{\text{los}}(r_{\perp})}{\bar{n}_{\text{los}}}. \quad (7.45)$$

I define the function  $\delta n = (n_{\text{los}}(r_{\perp})/\bar{n}_{\text{los}}) - 1$ . The multipoles of the tomographic effect correction can be written

$$\xi_{\text{v}\alpha,\ell}^s(r) = \int_{-1}^1 \left( \frac{1+2\ell}{2} \right) P_{\ell}(\mu) \xi_{\text{v}\alpha}(r) (1 + \delta n(r_{\perp})) d\mu. \quad (7.46)$$

Since we do not consider RSD, we can write  $\xi_{\text{v}\alpha}^s(\vec{r}) = \xi_{\text{v}\alpha}(r) + \xi_{\text{tomo}}^s(\vec{r})$ . As previously, I perform the coordinate transformation  $\mu^2 = 1 - x^2$ , so that one gets:

$$\xi_{\text{tomo},\ell} = \xi_{\text{v}\alpha}(r) \left( \frac{1+2\ell}{2} \right) \int_{-1}^1 \frac{xdx}{\sqrt{1-x^2}} P_{\ell}(\sqrt{1-x^2}) \delta n(xr). \quad (7.47)$$

I define the weighted averages of  $\delta n$  as follow:

$$\langle \delta n \rangle_i = \frac{\int n(xr) x^i / \sqrt{1-x^2}}{\int x^i / \sqrt{1-x^2}}. \quad (7.48)$$

As the previous model, I use the known integrals  $\int_0^1 x^i dx / \sqrt{1-x^2} = 1, 2/3, 8/15$  for  $i = 1, 3, 5$  respectively. The tomographic effect contributions for all multipoles are given in this second toy model by

$$\begin{aligned}
\xi_{\text{tomo},0}(r) &= \xi_{\text{v}\alpha}(r) \langle \delta n \rangle_1, \\
\xi_{\text{tomo},2}(r) &= \xi_{\text{v}\alpha}(r) (5 \langle \delta n \rangle_1 - \langle \delta n \rangle_3), \\
\xi_{\text{tomo},4}(r) &= \xi_{\text{v}\alpha}(r) (9 \langle \delta n \rangle_1 - 30 \langle \delta n \rangle_3 + 21 \langle \delta n \rangle_5).
\end{aligned}
\tag{7.49}$$

Predictions for this second toy model are a bit different from the first one. The radial shapes of tomographic multipoles scale like  $\xi_{\text{v}\alpha}(r)$  rather than its derivative. Furthermore, the predicted hexadecapole is very close to zero due to the cancellation of terms in Eq. 7.49, in the limit when  $\langle \delta n \rangle_i$  is independent of  $i$ . Overall, this second model seems to be less effective to understand the features seen in the mocks.

For both models developed in this section, I find that, to first order, the tomographic effect generates an additive correction  $\xi_{\text{tomo}}$  to the cross-correlation. Furthermore, after multipole expansion this correction factorizes  $\xi_{\text{tomo},\ell} = A_\ell f(r)$ , where both the  $A_\ell$  coefficients and the radial term  $f(r)$  are model-dependent.

### 7.3.3 Other tests and model summary

#### Disentangling the RSD and tomographic effect

Using to the mocks with and without RSD, it is possible to observe how RSD can be disentangled from the tomographic effect. The top panel of Fig. 7.15 shows the Ly $\alpha$   $\times$  Void cross-correlation as a function of  $\mu^2$ , in two different radial bins, of the *full11-RSD* and *full11-noRSD* mocks.

The observed dependence in  $\mu^2$  is not linear, as expected due to the impact of the tomographic effect. On the other hand, the bottom panel of Fig. 7.15 shows that the difference  $\xi_{\text{v}\alpha}^{\text{RSD}} - \xi_{\text{v}\alpha}^{\text{noRSD}}$  is better fit by a linear function of  $\mu^2$ . We interpret this feature by assuming that the contribution of tomographic effect to  $\xi_{\text{v}\alpha}$  is additive. This is what is predicted, at first order, by our toy models. When this additive contribution from the tomographic effect is removed, the measured slopes are then essentially due to the linear RSD effect. Consequently, the RSD contribution to the quadrupole can be recovered at least in a first approximation by subtracting the tomographic effect using mock without RSD.

#### Attempts of numerical corrections

I performed tests on the mocks to try to correct the tomographic effect directly, by modifying either the cross-correlation or the void catalog. These corrections use the  $\epsilon$  function as obtained from the random void catalog presented in Sec. 7.3.2.

The first method implemented is a modification of the cross-correlation calculation. I increase the separation of all Ly $\alpha$   $\times$  Void pairs such that  $r_{\perp,ij} = s_{\perp,ij} + \epsilon(s_{\perp,ij})$  where  $s_{\perp,ij}$  is computed using Eqn. 7.4. The second method I adopted is to correct the position of the void centers of my catalog. I use a barycenter scheme to displace void in the transverse direction by  $\epsilon$ .

I implemented these two methods and performed tests on the *full-noRSD* mock. The monopole increased for both techniques, indicating that the voids are indeed misplaced. However, the two corrections did not reduce the non-null quadrupole due to tomographic effect. It means that those simple corrections are not sufficient at this stage.

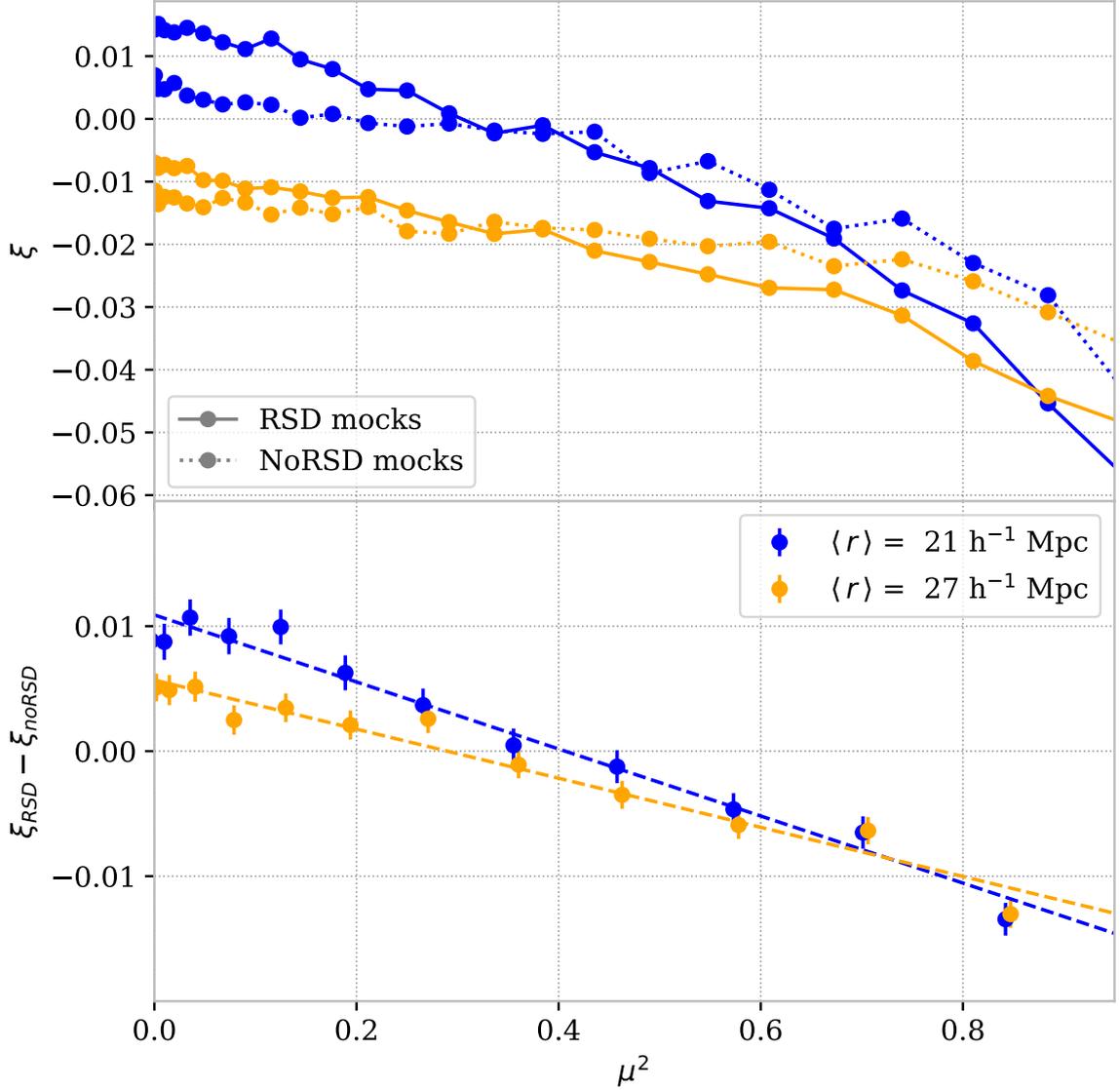


FIGURE 7.15 – Cross-correlations with  $A = (r, \mu)$  binning averaged on two radial bins  $\langle r \rangle = 21 \text{ h}^{-1} \cdot \text{Mpc}$  (blue) and  $27 \text{ h}^{-1} \cdot \text{Mpc}$  (yellow) of the average *full11-RSD* (continuous line) and *full11-noRSD* (dotted line) mocks. The cross-correlations are expressed as a function of  $\mu^2$ , considering both  $\mu < 0$  and  $\mu > 0$  bins. On the bottom panel, the difference between *full11-RSD* and *full11-noRSD* mocks is represented and the dashed lines show associated linear fits.

### Full model for the Ly $\alpha$ $\times$ Void cross-correlation

I model the RSD-induced quadrupole in  $\xi_{\nu\alpha}$ , starting from Eqn. 7.34. I must then correct the quadrupole due to the tomographic effect. Amplitude and shape measurements of the multipoles from the *full-noRSD* mock are in qualitative agreement with prediction of the toy models. However, due to their simplicity, I cannot rely quantitatively on those models as illustrated by the mismatch between the multipole ratios. To take into account the tomographic effect, I therefore rely on mocks, and on the approximation that the tomographic effect contribution to  $\xi_{\nu\alpha}$  is additive. Consequently, I subtract the quadrupole of the mocks without RSD:  $\xi_{\text{tomo},2} = \xi_{\nu\alpha,2}^{\text{noRSD}}$ .

As demonstrated by the toy models, the tomographic effect also modifies the monopole of the Ly $\alpha$   $\times$  Void cross-correlation monopole. For both models, we found that  $\xi_{\text{tom},0}(r)$  is proportional to  $\xi_{\text{tom},2}(r)$ , although the correction coefficient is different for the two models. Consequently, I decided to model  $\xi_{\text{tom},0}(r)$  by  $k \xi_{\text{va},2}^{\text{noRSD}}(r)$ , where  $k$  is a nuisance parameter. The corrected monopole therefore becomes  $\xi_{\text{va},0}(r) - k \xi_{\text{va},2}^{\text{noRSD}}(r)$ . Tests on mocks showed that this correction is important and cannot be neglected.

I do not take into account the instrumental and astrophysical effects discussed in Sec. 7.2.3, as their impact on the quadrupole is subdominant compared to both the tomographic effect and statistical uncertainties (see Fig. 7.11).

Finally, dropping the  $^s$  notation used in Eqn. 7.34 for clarity, the final model is

$$\xi_{\text{va},2}(r) - \xi_{\text{va},2}^{\text{noRSD}}(r) = \frac{2\beta}{3 + \beta} \times \left[ (\xi_{\text{va},0}(r) - k \xi_{\text{va},2}^{\text{noRSD}}(r)) - (\bar{\xi}_0(r) - k \bar{\xi}_{\text{va},2}^{\text{noRSD}}(r)) \right].$$

(7.50)

## 7.4 Lyman- $\alpha$ $\times$ Void cross-correlation in SDSS

### 7.4.1 Data measurement and comparison with mocks

Finally, let's turn to eBOSS Stripe 82 data. I used the *clean-xcorr* void catalog described in Sec. 7.1.1 on eBOSS data. I find 2906 voids located in the redshift range  $z = 2.1 - 3.2$ , with a median redshift 2.49. Void radii are in the range  $7 - 35 h^{-1} \cdot \text{Mpc}$ , with a median radius of  $13.0 h^{-1} \cdot \text{Mpc}$ . A fraction of 90 % of void radii are in the range  $[8.0 - 22.7] h^{-1} \cdot \text{Mpc}$ . The measured cross-correlation in the  $(r_{\parallel}, r_{\perp})$  plane, and its multipole expansion, is shown in both Fig. 7.16 and 7.17. Clearly, the comparison by eye with mocks shows that the measured quadrupole is a good match to mocks including RSD: we do observe the characteristic impact of velocity flows around voids.

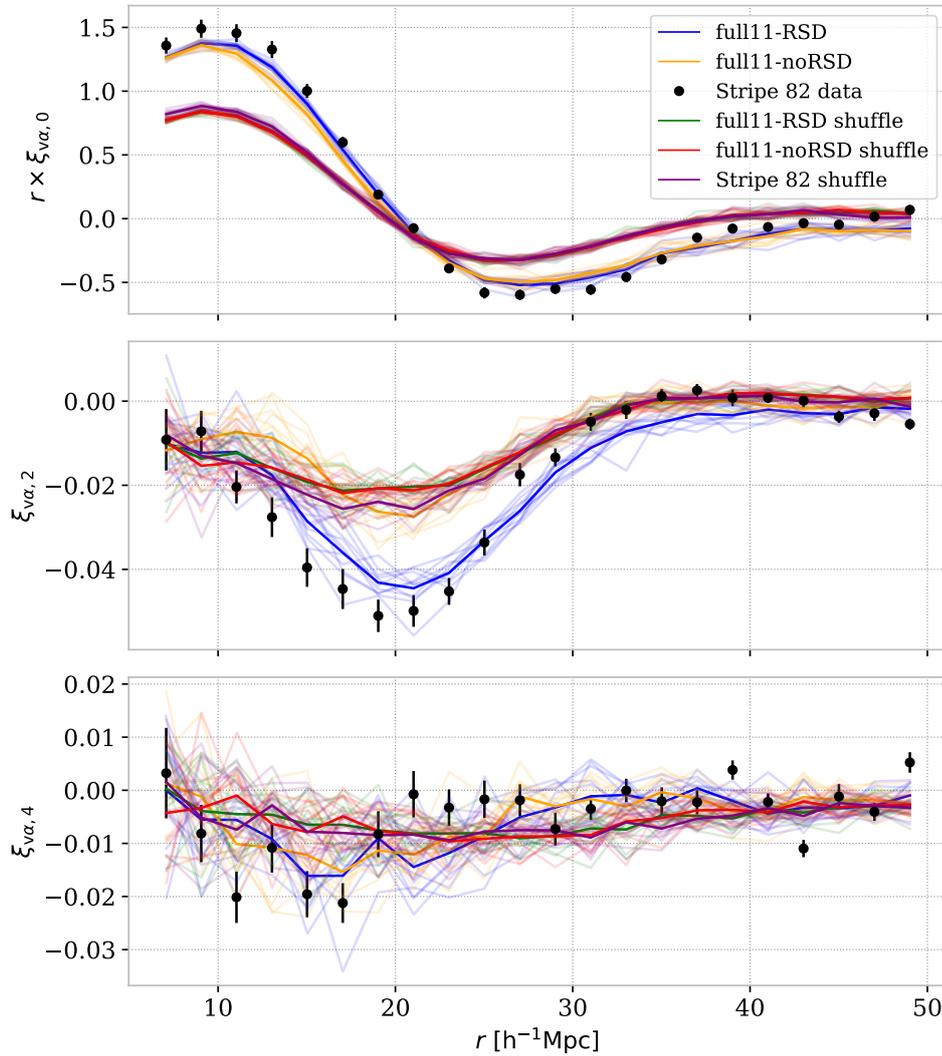


FIGURE 7.16 – Monopoles, quadrupoles and hexadecapoles of the Ly $\alpha$   $\times$  Void cross-correlation for eBOSS Stripe 82 data, *full11-RSD*, and *full11-noRSD* mocks. Multipoles of the shuffling of the same data and mocks are also reported. Thin curves represent individual mock or shuffle realizations, and their average is shown with thick curves.

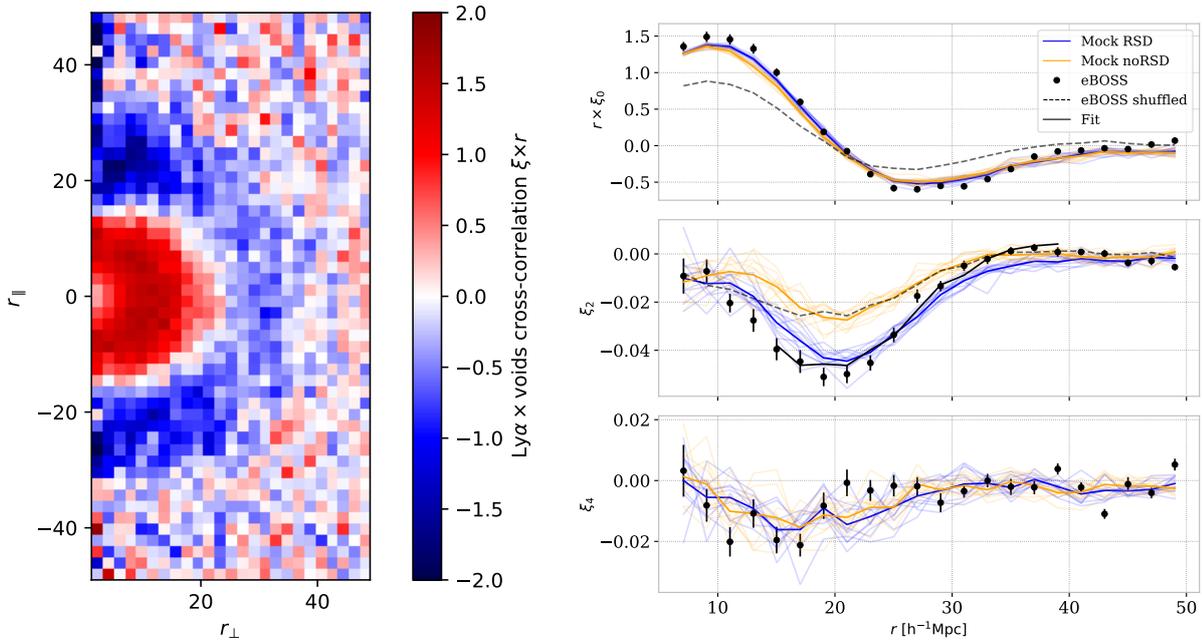


FIGURE 7.17 – (left) 2D representation of the Ly $\alpha$ ×Void cross-correlation in  $A = (r_{\parallel}, r_{\perp})$  binning of the eBOSS Stripe 82 data. (right) Monopoles, quadrupoles and hexadecapoles of the Ly $\alpha$ ×Void cross-correlation for eBOSS Stripe 82 data (black points), *full11-RSD* (blue), and *full11-noRSD* (orange) mocks. Thin curves represent individual mock realizations, and their average is shown with thick curves. Black dashed curves show the average monopole and quadrupole measured after shuffling eBOSS data. The continuous black curve shows the fit of the eBOSS quadrupole with Eqn. 7.50.

Similar to the **Systematics** mocks, the mid panel of Fig. 7.17 (right) demonstrates that the observed quadrupole has two main contributions, which are velocity flows and the tomographic effect. In this case, thanks to the multiple mock realizations, the separation between mocks including RSD or not is way more apparent. The quadrupole measured from eBOSS data is in agreement with *full11-RSD* mocks. Therefore, I observe the average flow of matter around voids from Ly $\alpha$  data, for radial separations  $\sim 15 - 30 h^{-1} \cdot \text{Mpc}$ .

I performed a shuffling test on both mocks and data. They are done with the same procedure as in Sec. 6.3.1. By shuffling the values of  $\delta_F$  between lines-of-sight, we keep the survey geometry while eliminating most physical Ly $\alpha$ ×Void and Ly $\alpha$ ×Ly $\alpha$  correlations including RSD. On the other hand, the tomographic effect is left unchanged. A void catalog and its Ly $\alpha$ ×Void cross-correlation are computed from these shuffles. The multipoles of the eBOSS data, *full11-noRSD* and *full11-noRSD* mocks, as well as the 11 shuffling realizations of the data and mocks are shown in Fig. 7.16. The shuffling reduces the magnitude of both the monopole and quadrupole (see Fig. 7.16). For all data and mock shuffles, we observe that the quadrupole is roughly equal to the one of *full11-noRSD* mocks, at least for  $r \gtrsim 15 h^{-1} \cdot \text{Mpc}$ . This is the case whether the RSD is included or not. This result confirms that the shuffling procedure removes the RSD signal while keeping the effects caused by the Ly $\alpha$  geometry and tomographic mapping. Consequently, the tomographic effect can also be removed from the data by subtracting the data shuffle quadrupole  $\xi_{\text{tomo},2} = \xi_{\text{v}\alpha,2}^{\text{DR16s}}$ . Even if this method is less theoretically motivated, it has the advantage of being fully data-driven. In any case, the equality between shuffles and *full11-noRSD* quadrupoles for  $r \gtrsim 15 h^{-1} \cdot \text{Mpc}$  indicates that we control the impact of the tomographic effect.

Statistical uncertainties for the data are estimated from the subsample covariance matrix of Eqn. 7.12. We checked that its diagonal terms agree with the measured variance from the eleven **Statistics** mocks within 10 %.

The monopole amplitude is slightly higher for the data than for mocks. This difference may result from residual differences between the measured and modeled noise distributions as pointed out in Chap. 6.

We observe a non-zero average hexadecapole in mocks, identical whether RSD is included or not. This result is in agreement with the linear flow model of Sec. 7.3.1 (which predicts a null RSD contribution to  $\xi_{v\alpha,4}$ ) and with the fact that the tomographic contribution to  $\xi_{v\alpha,l}$  is essentially additive. The hexadecapole in eBOSS data is consistent with mocks, but given the statistical uncertainties, it cannot be distinguished from zero for separations larger than approximately  $17 h^{-1}\cdot\text{Mpc}$ .

## 7.4.2 Model fitting

Thanks to the model developments presented in Sec. 7.3, one can play the game of measuring the void-Ly $\alpha$  RSD parameter  $\beta$ . I also fit the shape of the measured monopole and compare it to the Galaxy  $\times$  Void cross-correlation studies.

### Void profile

As shown in the article [24], the void monopole in Galaxy  $\times$  Void can be modeled by the following universal profile

$$\xi_{\text{raw},0} = \delta_c \frac{1 - (r/r_s)^\alpha}{1 - (r/r_v)^\beta}. \quad (7.51)$$

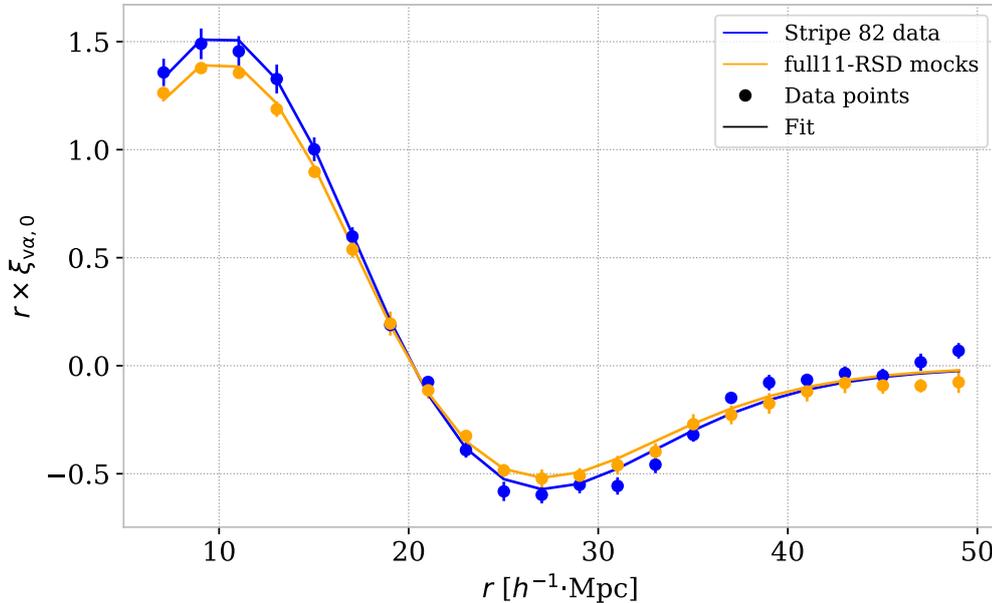


FIGURE 7.18 – Monopoles of the eBOSS Stripe 82 data and *full11-RSD* average mock, both represented by points. Plain curves represent the fit by a Gaussian smoothed void profile as given by Eqn. 7.51. The resulting values of this fit are reported in Tab. 7.2.

	$r_v$ [ $h^{-1}\cdot\text{Mpc}$ ]	$\delta_c$	$d_{\text{filter}}$ [ $h^{-1}\cdot\text{Mpc}$ ]
Stripe 82 data	17.0	0.47	7.6
<i>full11-RSD</i>	16.8	0.44	7.6
<i>full11-noRSD</i>	16.5	0.44	7.5

TABLE 7.2 – Results of the monopole fit of the eBOSS Stripe 82 data and **Statistics** average mocks using the Gaussian smoothed void profile.

Since my voids are found from a smoothed map, their positions are blurred by this smoothing. Consequently, I convolved this profile with a Gaussian function  $\xi_{\text{model},0} = \xi_{\text{raw},0} \otimes \text{Gauss}(d_{\text{filter}})$ . As in [24], I fix the values  $\alpha = 2$ ,  $\beta = 9.6$  and  $r_s = 0.82 r_v$ . I performed this fit for the Stripe 82 data and **Statistics** mocks. The results are shown in Fig. 7.18 and the values reported in Tab. 7.2.

As shown in the figure, the void profile is well fitted by the model, almost identically in the case of mocks or data.

### RSD parameter fit

The fit of the Ly $\alpha$  $\times$ Void cross-correlation quadrupole is done using the model given by Eqn. 7.50. The result of this fit is represented in Fig. 7.17 and with more details in Fig. 7.19.

As shown in the previous section, the tomographic quadrupole  $\xi_{\text{tomo},2}$  can be estimated using the average quadrupole either from *full11-noRSD* mocks or from data shuffle. Here, I choose to use the data shuffle to obtain a "mock-independent" fit.

In Eqn. 7.50, there are two free parameters,  $\beta$  and  $k$ . The fit is realized for  $15 \leq r \leq 40 h^{-1}\cdot\text{Mpc}$ , and I use statistical errors from the covariance matrix of  $\xi_{v\alpha,2}$ . The fitting procedure yields:

$$\begin{aligned} k &= -0.9 \pm 0.3, \\ \beta &= 0.52 \pm 0.05. \end{aligned} \tag{7.52}$$

This result corresponds to a detection of the RSD effect formally at  $10\sigma$  statistical significance. The same fit procedure applied to the average of *full11-RSD* mocks consistently yields  $\beta = 0.58 \pm 0.06$ .

I checked that the inferred value for  $\beta$  is largely independent of  $k$ . Forcing  $k = 0$  or varying its prior still yield  $\beta = 0.50 - 0.52$  in the data fit. The fit executed with  $k$  fixed to zero is reported in Fig. 7.19. Likewise, Using the average quadrupole from *full11-noRSD* mocks as an estimator for  $\xi_{\text{tomo},2}$ , I find  $\beta = 0.54 \pm 0.05$  which indicates this change is not significant.

As illustrated by Fig. 6.13, voids with radii  $< 15 h^{-1}\cdot\text{Mpc}$  are contaminated by noise fluctuations. Using mock data, I performed the same measurement after selecting voids with radius  $> 15 h^{-1}\cdot\text{Mpc}$ , keeping 40% of the void statistics. I observed qualitatively identical patterns in the multipoles for *full11* mocks and data. The characteristic scales of the Ly $\alpha$  $\times$ Void cross-correlation are increased. For example, the crossing of the monopole with the x-axis is shifted from 20 to 24  $h^{-1}\cdot\text{Mpc}$ . In particular, the relative amplitudes of quadrupoles from *full11-RSD* and *full11-noRSD* mocks are unchanged. On the other hand, the measured statistical error on the quadrupole is increased by 60%. This justifies the choice to use smaller voids as well for this specific study: contamination by spurious voids are more than balanced by the additional statistics.

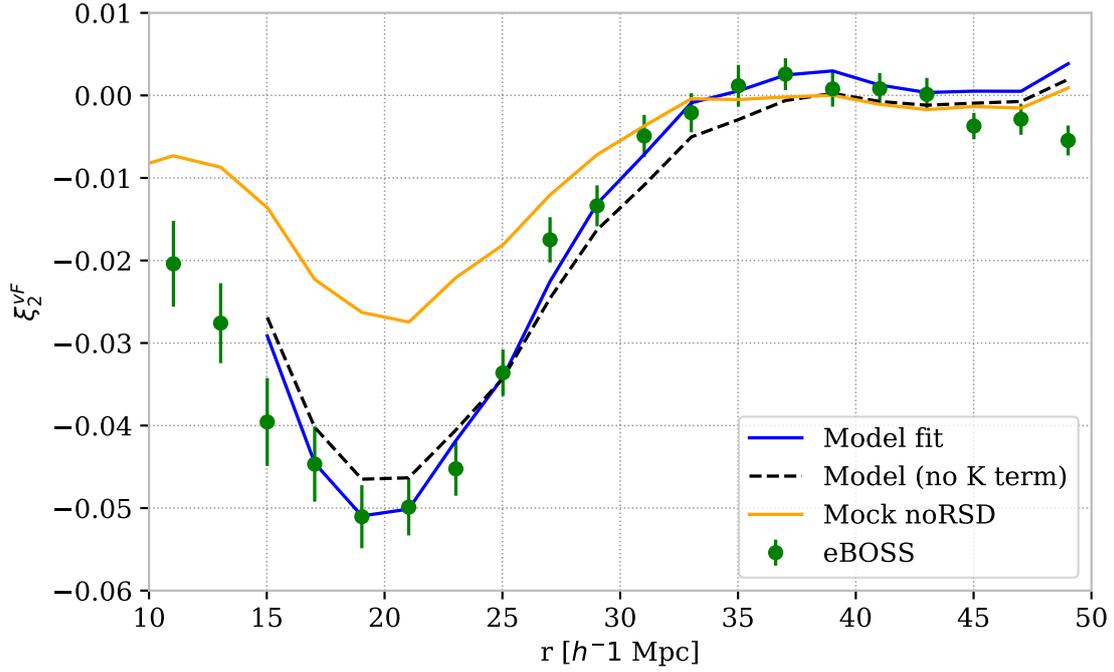


FIGURE 7.19 – Quadrupole of the eBOSS Stripe 82 data (green points) and the *full11-noRSD* average mock (yellow). The data fit using the model of Eqn. 7.50 is represented in blue. The dashed curve shows the fit obtained when  $k$  is fixed to zero. In both cases, the average quadrupole of the *full11-noRSD* mocks is subtracted in the fit to remove the tomographic effect.

The measured void-Ly $\alpha$  RSD parameter,  $\beta = 0.52 \pm 0.05$ , is smaller than a similar parameter inferred from the large-scale eBOSS Ly $\alpha$  auto-correlation:  $\beta_{\text{auto}} = 1.66$  [16]. This difference between both observational quantities is reproduced when analysing mock data. It suggests that the linear bias coefficient for the Ly $\alpha$  forest in the vicinity of voids (the void-Ly $\alpha$  bias  $b$  defined by Eqn. 7.30) differs from the Ly $\alpha$  bias  $b_\alpha$  defined in Eqn. 2.29. A complete study of the void-Ly $\alpha$  and velocity biases from hydrodynamical simulations is required to interpret the  $\beta$  measurement fully.

## Bibliography

- [1] EBOSS collaboration, *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological Implications from two Decades of Spectroscopic Surveys at the Apache Point observatory*, *Physical Review D* **103** (2021) 083533.
- [2] S. Nadathur and S. Hotchkiss, *The nature of voids - I. Watershed void finders and their connection with theoretical models*, *Monthly Notices of the Royal Astronomical Society* **454** (2015) 2228.
- [3] S. Nadathur and S. Hotchkiss, *The nature of voids: II. Tracing underdensities with biased galaxies*, *Monthly Notices of the Royal Astronomical Society* **454** (2015) 889.
- [4] Y.-C. Cai, A. Taylor, J. A. Peacock and N. Padilla, *Redshift-space distortions around voids*, *Monthly Notices of the Royal Astronomical Society* **462** (2016) 2465.
- [5] N. Hamaus, M.-C. Cousinou, A. Pisani, M. Aubert, S. Escoffier and J. Weller, *Multipole analysis of redshift-space distortions around cosmic voids*, *Journal of Cosmology and Astroparticle Physics* **2017** (2017) 014.
- [6] S. Nadathur, P. M. Carter, W. J. Percival, H. A. Winther and J. Bautista, *Beyond BAO: improving cosmological constraints from BOSS with measurement of the void-galaxy cross-correlation*, *Physical Review D* **100** (2019) 023504.
- [7] C. Alcock and B. Paczyński, *An evolution free test for non-zero cosmological constant*, *Nature* **281** (1979) 358.
- [8] G. Lavaux and B. D. Wandelt, *Precision cosmography with stacked voids*, *The Astrophysical Journal* **754** (2012) 109.
- [9] V. Demchenko, Y.-C. Cai, C. Heymans and J. A. Peacock, *Testing the spherical evolution of cosmic voids*, *Monthly Notices of the Royal Astronomical Society* **463** (2016) 512.
- [10] C. Ravoux, E. Armengaud, J. Bautista, J.-M. L. Goff, N. Palanque-Delabrouille, J. Rich et al., *First measurement of the correlation between cosmic voids and the Lyman- $\alpha$  forest*, *arXiv:2203.11045 [astro-ph]* (2022) .
- [11] A. Slosar, V. Iršič, D. Kirkby, S. Bailey, N. G. Busca, T. Delubac et al., *Measurement of Baryon Acoustic Oscillations in the Lyman-alpha Forest Fluctuations in BOSS Data Release 9*, *Journal of Cosmology and Astroparticle Physics* **2013** (2013) 026.
- [12] A. Font-Ribera, D. Kirkby, N. Busca, J. Miralda-Escudé, N. P. Ross, A. Slosar et al., *Quasar-Lyman  $\alpha$  Forest Cross-Correlation from BOSS DR11 : Baryon Acoustic Oscillations*, *Journal of Cosmology and Astroparticle Physics* **2014** (2014) 027.
- [13] T. Delubac, J. E. Bautista, N. G. Busca, J. Rich, D. Kirkby, S. Bailey et al., *Baryon Acoustic Oscillations in the Ly $\alpha$  forest of BOSS DR11 quasars*, *Astronomy & Astrophysics* **574** (2015) A59.
- [14] J. E. Bautista, N. G. Busca, J. Guy, J. Rich, M. Blomqvist, H. d. M. d. Bourboux et al., *Measurement of BAO correlations at  $z=2.3$  with SDSS DR12 Ly $\alpha$ -Forests*, *Astronomy & Astrophysics* **603** (2017) A12.

- [15] H. d. M. d. Bourboux, J.-M. L. Goff, M. Blomqvist, N. G. Busca, J. Guy, J. Rich et al., *Baryon acoustic oscillations from the complete SDSS-III Ly $\alpha$ -quasar cross-correlation function at  $z=2.4$* , *Astronomy & Astrophysics* **608** (2017) A130.
- [16] H. d. M. d. Bourboux, J. Rich, A. Font-Ribera, V. d. S. Agathe, J. Farr, T. Etourneau et al., *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Baryon acoustic oscillations with Lyman- $\alpha$  forests*, *The Astrophysical Journal* **901** (2020) 153.
- [17] H. du Mas des Bourboux, J. Rich, A. Font-Ribera, V. de Sainte Agathe, J. Farr, T. Etourneau et al., *picca: Package for Igm Cosmological-Correlations Analyses*, *Astrophysics Source Code Library* (2021) ascl:2106.018.
- [18] P. McDonald, U. Seljak, S. Burles, D. J. Schlegel, D. H. Weinberg, D. Shih et al., *The Lyman-alpha Forest Power Spectrum from the Sloan Digital Sky Survey*, *The Astrophysical Journal Supplement Series* **163** (2006) 80.
- [19] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke et al., *HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere*, *The Astrophysical Journal* **622** (2005) 759.
- [20] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau et al., *SciPy 1.0: fundamental algorithms for scientific computing in Python*, *Nature Methods* **17** (2020) 261.
- [21] U. Seljak, *Bias, redshift space distortions and primordial nongaussianity of nonlinear transformations: application to Lyman alpha forest*, *Journal of Cosmology and Astroparticle Physics* **2012** (2012) 004.
- [22] P. M. Sutter, P. Elahi, B. Falck, J. Onions, N. Hamaus, A. Knebe et al., *The life and death of cosmic voids*, *Monthly Notices of the Royal Astronomical Society* **445** (2014) 1235.
- [23] G. Pollina, N. Hamaus, K. Dolag, J. Weller, M. Baldi and L. Moscardini, *On the linearity of tracer bias around voids*, *Monthly Notices of the Royal Astronomical Society* **469** (2017) 787.
- [24] N. Hamaus, P. M. Sutter and B. D. Wandelt, *Universal Density Profile for Cosmic Voids*, *Physical Review Letters* **112** (2014) 251302.



---

# Conclusion and prospects

---

*“Toute conquête, chaque pas en avant dans le domaine de la connaissance a son origine dans le courage, dans la dureté à l’égard de soi-même, dans la propreté vis-à-vis de soi-même.”*

– Friedrich Nietzsche, *Ecce Homo*, 1888

## General conclusion

Our view of the Universe is made arduous by the still unknown nature of its components and practical limitations in terms of observations. Even if the  $\Lambda$ CDM model presented in Chap. 1 currently offers us a globally unified description of the Universe, it remains imperfect because of tensions between observations and still unexplained phenomena. The research effort in cosmology consists of modeling, simulating, and maximizing the amount of information extracted from observations. These different methods are the tools at our disposal to unravel the Universe’s history and structure. I tried to contribute with this thesis to this effort with the study of the Ly $\alpha$  forest, detailed in Chap. 2, both by improving its one-dimensional analysis and by designing methods to exploit it in three dimensions. To this end, I have worked on the data provided by the SDSS and DESI telescopes presented in Chap. 4. Doing this research work within the CEA cosmology group and the SDSS and DESI collaborations was a very enriching experience in a supportive and stimulating work environment.

In Chap. 3, I presented the simulation tools useful for the analysis of the Ly $\alpha$  forest. In particular, I detailed the approximated log-normal mocks used to mimic the SDSS and DESI data. These synthetic data were used to model the impact of Ly $\alpha$  systematics such as Continuum fitting, noise, BAL, HCD, metals, and sky emission lines. For cosmological inferences, modeling the one-dimensional Ly $\alpha$  power spectrum ( $P_{1D,\alpha}$ ) requires simulating highly non-linear processes with hydrodynamical simulations of the intergalactic medium. I contributed to the creation of

a Nyx Gaussian processes emulator used for the cosmological interpretation of  $P_{1D,\alpha}$  [1]. This emulator uses an adapted sampling of the cosmological parameter space to reduce the number of simulations needed for cosmological inferences. In parallel, I co-led a project on the Jean Zay supercomputer to create a set of large-scale hydrodynamical simulations (up to  $500 h^{-1}\cdot\text{Mpc}$ ) using an accelerated Nyx GPU version. We used this set to improve the prediction from simulations of the three-dimensional Ly $\alpha$  power spectrum ( $P_{3D,\alpha}$ ). This work is currently ongoing and will lead to a publication.

A major part of this thesis is dedicated to the FFT measurement of the one-dimensional Ly $\alpha$  power spectrum ( $P_{1D,\alpha}$ ) from the first DESI data, presented in Chap. 5. The DESI spectral resolution is approximately two times better than SDSS. Consequently, a measurement of  $P_{1D,\alpha}$  is of high scientific interest to probe the small scales of the intergalactic medium. This measurement is sensitive to instrumental noise so that I exchanged with the DESI pipeline team to verify the noise estimation on several successive intermediate data releases. I studied instrumental and theoretical systematics associated with this analysis and transposed them on the  $P_{1D,\alpha}$  measurement. This ongoing project will lead to an article using DESI SV+DA0.2 data.

From the BOSS and eBOSS quasar spectra measured within the specific Stripe 82 field, I constructed in Chap. 6 a near-Gpc<sup>3</sup> volume 3D tomographic map of the Ly $\alpha$  contrast [2]. Its  $13 h^{-1}\cdot\text{Mpc}$  resolution is such that it directly maps the large-scale matter distribution, both baryonic and dark. This study constitutes the most extensive Ly $\alpha$  tomographic map thus far and uniquely represents matter fluctuations at high redshifts ( $z > 2$ ). From log-normal mocks, I evaluated that this tomographic map is correlated with a  $-34\%$  coefficient to the underlying matter fluctuations. I developed a multi-threaded void finder and presented a search for large voids over the volume of the map. The mean absorption profile around voids indicated that their geometry is not isotropic. Stacking the tomographic map around quasars yields a clear cross-correlation signal. The map is also well adapted to protocluster searches with a radius larger than  $7 h^{-1}\cdot\text{Mpc}$ ; I provide a list of candidate protoclusters identified by a watershed algorithm. I also co-led a secondary target program that aimed to observe Lyman-break galaxies (LBGs) with DESI. The first results obtained in the COSMOS field provide encouraging results for the use of LBGs in future surveys.

Voids are interesting to probe the effect of cosmic expansion on structure growth as the physics describing them is more linear than dense halos. From the eBOSS tomographic map, I carried out in Chap. 7, a measurement of the Ly $\alpha$ ×Void cross-correlation at a median redshift  $z = 2.49$  [3]. The purpose of the study was to use voids and the entire shape of their correlation to observe redshift space distortions (RSD) in the Ly $\alpha$  redshift range. I built a linear model by adapting void-galaxy methods to the Ly $\alpha$  forest framework and expanded it to include effects related to Ly $\alpha$  tomography. I used this model to fit a void RSD parameter  $\beta = 0.52 \pm 0.05$ . It is the first time that voids detected from the Ly $\alpha$  forest are used as a cosmological probe and the first observation of large-scale velocity flows around voids at such a high redshift. This work paves the way toward growth-rate measurements at redshifts currently inaccessible to galaxy surveys.

# Prospects

## Method and modeling improvements

The emulator presented in Sec. 3.2.3 appears to be a promising technique to reduce the CPU time needed to run simulations. It could be extended by considering additional effects, astrophysical and cosmological. For example, neutrinos can be added to the emulator by incorporating them as an additional species with the approximated methods presented in Sec. 3.2. Warm dark matter and other exotic dark matter models can also be easily included as long as their impact is limited to correcting the initial power spectrum. One of the main limitations of the Ly $\alpha$  probe is the systematics associated with the physics of the intergalactic medium. The temperature, velocity, and density of baryons are influenced by astrophysical effects such as supernovae, active galactic nuclei feedback, and inhomogeneous re-ionization. Adding a parameterized impact of these effects to an emulator would significantly improve the control of such systematics. Emulators can also constrain modified gravity models such as f(R) cosmologies. Based on existing theoretical works [4], I think coupling this f(R) emulator with precise enough  $P_{1D,\alpha}$  measurement will probably bring constraints on this scenario.

The  $P_{1D,\alpha}$  measurement can be improved by using different estimations, such as the maximum likelihood quadratic method, already used for high-resolution spectra [5, 6]. A  $P_{3D,\alpha}$  measurement has never been performed because of previous surveys' low quasar angular density. If a large portion of the DESI footprint has a homogeneous density of Ly $\alpha$  quasars, a  $P_{3D,\alpha}$  measurement can be considered using a cross power-spectrum method [7] or an improvement of the Ly $\alpha$  tomographic map presented in Chap. 6. Such a measurement will combine the small-scale information along the lines-of-sight together with transverse correlations. Modeling the  $P_{3D,\alpha}$  requires the development of extensive hydrodynamical simulations like the one we developed in Sec. 3.2.4. With an adapted parameterization of the simulations,  $P_{3D,\alpha}$  could be used to improve cosmological parameter inferences. In addition, our simulations can be exploited to improve current Ly $\alpha$  auto-correlation models and use its entire shape as suggested in [8].

Ly $\alpha$  tomography applications are promising as they can potentially recover higher-order statistics with respect to standard Ly $\alpha$  analysis. The Wiener filter technique, used in Chap. 6, reproduces the global Gaussian fluctuations of Ly $\alpha$  absorption, which is appropriate for void finding. However, more advanced methods are needed to directly use the statistical properties of the regions outside the lines-of-sight. A concrete example is the adaptation of inpainting methods, which have been already applied to CMB [9, 10]. By associating it with a Wiener filter, the resulting map will contain global variations and a more accurate representation of over-densities. More advanced algorithms can be used, in particular following the forward modeling approach [11, 12, 13, 14]. They are expected to make it possible to directly estimate the underlying matter field from an input Ly $\alpha$  forest sample. It requires the assumption of a model like FGPA to determine the Ly $\alpha$  quantities. A possible improvement would be to replace FGPA by deep learning inference of hydrodynamical fields such as those developed by [15, 16]. Using a forward modeling technique on a wide survey would provide a tomographic map that could potentially be used to measure high-order statistics accurately. Increasing the source density will also improve tomographic mapping. For example, using LBG as additional tracers as we tested with our secondary target program will greatly improve the line-of-sight density. Finally, Ly $\alpha$  tomographic map could be used to improve CMB lensing cross-correlation studies as pioneered

by [17]. One can compute the cross-correlation not only with individual one-dimensional power spectra as in [17] but with an integration of the Ly $\alpha$  tomographic map along the radial direction over a given redshift range. This method would allow cross-correlating two angular fields and linking the matter density fluctuations from different probes with their associated systematics.

For now, most of the applications using tomographic mapping are proto-cluster identifications [18, 19, 20]. My work on the use of voids presented in Chap. 7 traces a promising path for void correlations at redshift higher than probed by galaxies [21, 22]. Improving map-making will reduce the tomographic effect and decrease the contamination by spurious voids. The void finder can also be improved, notably by searching for non-spherical voids. Another approach considering voids as critical points of the cosmic web [23, 24] can potentially be used to improve this study. Accessing the full potential of the Ly $\alpha$  $\times$ Void cross-correlation requires a better understanding of void physics with large-scale hydrodynamical simulations like the one developed in Chap. 3. This improved modeling could help decouple RSD and Alcock-Paczynski (AP) effects on voids. Furthermore, simulations showed that voids could be used to improve neutrino mass constraints [25].

### Prospect for ongoing and future surveys

The data from DESI [26, 27], but also WEAVE-QSO [28] surveys will improve all the measurements developed in this thesis. Regarding the P1D measurement, the full DESI survey will bring enough statistics to achieve a percent precision measurement over a large redshift range. Achieving this precision will require an unprecedented level of control of the systematics, especially at small scales. With both DESI and WEAVE-QSO surveys, average line-of-sight separations down to  $7 h^{-1} \cdot \text{Mpc}$  can be achieved and lead to more extended and refined tomographic maps. Additionally, the increase in the angular density of quasar spectra observed by DESI and WEAVE-QSO will also drastically increase the voids statistics. Consequently, the cross-correlation measurement we performed will be greatly improved.

Finally, recent and future surveys will mostly bring full-sky measurements for large-scale structures (eBOSS, DESI, Euclid), supernovae (ZTF, Rubin Observatory), and CMB (Planck, ACT/SPT, CMB-S4). All these surveys will offer unprecedented opportunities for cosmic probe combinations. I am convinced that maximizing the scientific potential of future experiments can only be done by combining those probes.

## Bibliography

- [1] M. Walther, E. Armengaud, C. Ravoux, N. Palanque-Delabrouille, C. Yèche and Z. Lukić, *Simulating intergalactic gas for DESI-like small scale Lyman $\alpha$  forest observations*, *Journal of Cosmology and Astroparticle Physics* **2021** (2021) 059.
- [2] C. Ravoux, E. Armengaud, M. Walther, T. Etourneau, D. Pomarède, N. Palanque-Delabrouille et al., *A tomographic map of the large-scale matter distribution using the eBOSS Stripe 82 Ly- $\alpha$  forest*, *Journal of Cosmology and Astroparticle Physics* **2020** (2020) 010.
- [3] C. Ravoux, E. Armengaud, J. Bautista, J.-M. L. Goff, N. Palanque-Delabrouille, J. Rich et al., *First measurement of the correlation between cosmic voids and the Lyman- $\alpha$  forest*, *arXiv:2203.11045 [astro-ph]* (2022) .
- [4] P. Brax and P. Valageas, *Lyman-alpha power spectrum as a probe of modified gravity*, *Journal of Cosmology and Astroparticle Physics* **2019** (2019) 049.
- [5] N. G. Karaçaylı, A. Font-Ribera and N. Padmanabhan, *Optimal 1D Ly- $\alpha$  Forest Power Spectrum Estimation I: DESI-Lite Spectra*, *Monthly Notices of the Royal Astronomical Society* **497** (2020) 4742.
- [6] N. G. Karaçaylı, N. Padmanabhan, A. Font-Ribera, V. Iršič, M. Walther, D. Brooks et al., *Optimal 1D Ly  $\alpha$  forest power spectrum estimation – II. KODIAQ, SQUAD, and XQ-100*, *Monthly Notices of the Royal Astronomical Society* **509** (2022) 2842.
- [7] A. Font-Ribera, P. McDonald and A. Slosar, *How to estimate the 3D power spectrum of the Lyman- $\alpha$  forest*, *Journal of Cosmology and Astroparticle Physics* **2018** (2018) 003.
- [8] A. Cuceu, A. Font-Ribera, B. Joachimi and S. Nadathur, *Cosmology beyond BAO from the 3D distribution of the Lyman- $\alpha$  forest*, *arXiv:2103.14075 [astro-ph]* (2021) .
- [9] J.-L. Starck, M. J. Fadili and A. Rassat, *Low- $l$  CMB Analysis and inpainting*, *Astronomy & Astrophysics* **550** (2013) A15.
- [10] G. Montefalcone, M. H. Abitbol, D. Kodwani and R. D. P. Grumitt, *Inpainting CMB maps using Partial Convolutional Neural Networks*, *Journal of Cosmology and Astroparticle Physics* **2021** (2021) 055.
- [11] B. Horowitz, K.-G. Lee, M. White, A. Krolewski and M. Ata, *TARDIS Paper I: A Constrained Reconstruction Approach to Modeling the  $z\sim 2.5$  Cosmic Web Probed by Lyman-alpha Forest Tomography*, *The Astrophysical Journal* **887** (2019) 61.
- [12] B. Horowitz, B. Zhang, K.-G. Lee and R. Kooistra, *TARDIS Paper II: Synergistic Density Reconstruction from Lyman-alpha Forest and Spectroscopic Galaxy Surveys with Applications to Protoclusters and the Cosmic Web*, *The Astrophysical Journal* **906** (2021) 110.
- [13] N. Porqueres, J. Jasche, G. Lavaux and T. Enßlin, *Inferring high redshift large-scale structure dynamics from the Lyman-alpha forest*, *Astronomy & Astrophysics* **630** (2019) A151.

- [14] N. Porqueres, O. Hahn, J. Jasche and G. Lavaux, *A hierarchical field-level inference approach to reconstruction from sparse Lyman- $\alpha$  forest data*, *arXiv:2005.12928 [astro-ph]* (2020) .
- [15] P. Harrington, M. Mustafa, M. Dornfest, B. Horowitz and Z. Lukić, *Fast, high-fidelity Lyman  $\alpha$  forests with convolutional neural networks*, *arXiv:2106.12662 [astro-ph]* (2021) .
- [16] F. Sinigaglia, F.-S. Kitaura, A. Balaguera-Antolínez, I. Shimizu, K. Nagamine, M. Sánchez-Benavente et al., *Mapping Lyman-alpha forest three-dimensional large scale structure in real and redshift space*, *arXiv:2107.07917 [astro-ph]* (2022) .
- [17] C. Doux, E. Schaan, E. Aubourg, K. Ganga, K.-G. Lee, D. N. Spergel et al., *First Detection of Cosmic Microwave Background Lensing and Lyman- $\alpha$  Forest Bispectrum*, *Physical Review D* **94** (2016) 103506.
- [18] K.-G. Lee, J. F. Hennawi, M. White, J. X. Prochaska, A. Font-Ribera, D. J. Schlegel et al., *Shadow of a Colossus: A  $z=2.45$  Galaxy Protocluster Detected in 3D Ly- $\alpha$  Forest Tomographic Mapping of the COSMOS Field*, *The Astrophysical Journal* **817** (2016) 160.
- [19] S. Mukae, M. Ouchi, Z. Cai, K.-G. Lee, J. X. Prochaska, S. Cantalupo et al., *3D Distribution Map of HI Gas and Galaxies Around an Enormous Ly $\alpha$  Nebula and Three QSOs at  $z=2.3$  Revealed by the HI Tomographic Mapping Technique*, *The Astrophysical Journal* **896** (2020) 45.
- [20] B. Horowitz, K.-G. Lee, M. Ata, T. Müller, A. Krolewski, J. X. Prochaska et al., *Second Data Release of the COSMOS Lyman-alpha Mapping and Tomographic Observation: The First 3D Maps of the Large-Scale Cosmic Web at  $2.05 < z < 2.55$* , *arXiv:2109.09660 [astro-ph]* (2021) .
- [21] N. Hamaus, M.-C. Cousinou, A. Pisani, M. Aubert, S. Escoffier and J. Weller, *Multipole analysis of redshift-space distortions around cosmic voids*, *Journal of Cosmology and Astroparticle Physics* **2017** (2017) 014.
- [22] S. Nadathur, P. M. Carter, W. J. Percival, H. A. Winther and J. Bautista, *Beyond BAO: improving cosmological constraints from BOSS with measurement of the void-galaxy cross-correlation*, *Physical Review D* **100** (2019) 023504.
- [23] J. Shim, S. Codis, C. Pichon, D. Pogosyan and C. Cadiou, *The clustering of critical points in the evolving cosmic web*, *Monthly Notices of the Royal Astronomical Society* **502** (2021) 3885.
- [24] K. Kraljic, C. Laigle, C. Pichon, S. Peirani, S. Codis, J. Shim et al., *Forecasts for WEAVE-QSO: 3D clustering and connectivity of critical points with Lyman- $\alpha$  tomography*, *arXiv:2201.02606 [astro-ph]* (2022) .
- [25] E. Massara, F. Villaescusa-Navarro, M. Viel and P. M. Sutter, *Voids in massive neutrino cosmologies*, *arXiv:1506.03088 [astro-ph]* (2015) .
- [26] DESI collaboration, *The DESI Experiment Part I: Science, Targeting, and Survey Design*, *arXiv:1611.00036 [astro-ph]* (2016) .

- 
- [27] DESI collaboration, *The DESI Experiment Part II: Instrument Design*, *arXiv:1611.00037 [astro-ph]* (2016) .
- [28] M. M. Pieri, S. Bonoli, J. Chaves-Montero, I. Paris, M. Fumagalli, J. S. Bolton et al., *WEAVE-QSO: A Massive Intergalactic Medium Survey for the William Herschel Telescope*, *arXiv:1611.09388 [astro-ph]* (2016) .



---

## Résumé substantiel

---

La Cosmologie a pour but de décrire l'évolution et la structure de l'Univers. Au cours de l'histoire, de nombreux modèles décrivant l'Univers ont été développés. Des descriptions religieuses aux théories actuelles, notre vision de l'Univers et notre place à l'intérieur de celui-ci ont été profondément modifiées.

Cette thèse est le résultat de mes efforts pour améliorer la compréhension de notre vaste Univers. J'ai réalisé ce travail dans le Département de Physique des particules (DPhP) au Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA), entre 2019 et 2022. L'objectif de ce résumé est d'introduire les concepts fondamentaux utilisés dans cette thèse et d'exposer mes contributions.

## I Le modèle standard de la Cosmologie

Lorsque l'on considère l'Univers dans sa globalité, sa dynamique est dictée par la gravitation. La théorie la mieux éprouvée à ce jour pour décrire la gravité est la Relativité Générale développée au début du XX<sup>ième</sup> siècle.

La description globale de l'Univers se base sur le **principe cosmologique** qui stipule qu'aux grandes échelles, l'Univers est uniforme et isotrope. Mathématiquement, on peut montrer que l'espace-temps de l'Univers peut être décrit par la **métrique de Friedmann-Lemaître-Robertson-Walker (FLRW)** :

$$ds^2 = -dt^2 + a^2(t) \left[ \frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2) \right]. \quad (7.53)$$

Cette métrique définit la géométrie de l'Univers dans un espace à trois dimensions spatiales  $(r, \theta, \phi)$  et une temporelle  $t$ . En utilisant la Relativité Générale, il est possible de déduire la dynamique de l'Univers à partir de cette métrique et de sa composition. La valeur de la **courbure**  $K$  définit la géométrie de l'Univers (sphérique, plat ou hyperbolique). Le facteur d'échelle  $a(t)$  correspond à la taille normalisée de l'Univers et donne son expansion.

Grâce aux observations réalisées depuis le début du XX<sup>ième</sup> siècle, et en particulier à celles d'Edwin Hubble qui a déterminé que toutes les galaxies lointaines s'éloignent de nous, nous savons que l'Univers est en expansion. Lorsque seule cette expansion est prise en compte, il est possible de relier la vitesse d'éloignement des galaxies  $v$  à leur distance  $d$  par la constante de Hubble  $H_0$ , tel que  $v = H_0 d$ . On caractérise également la distance d'un objet grâce au **redshift**  $z$ , le rapport entre la longueur d'onde observée  $\lambda_0$  et la longueur d'onde d'émission  $\lambda_e$  d'un objet. En considérant que cette vitesse est entièrement dictée par l'expansion de l'Univers, on obtient la relation

$$1 + z = \frac{\lambda_0}{\lambda_e} = \frac{a(t=0)}{a(t_e)}. \quad (7.54)$$

Le redshift est utilisé pour localiser une source dans "l'espace des redshifts", qui est un espace pratique adapté aux observations spectroscopiques.

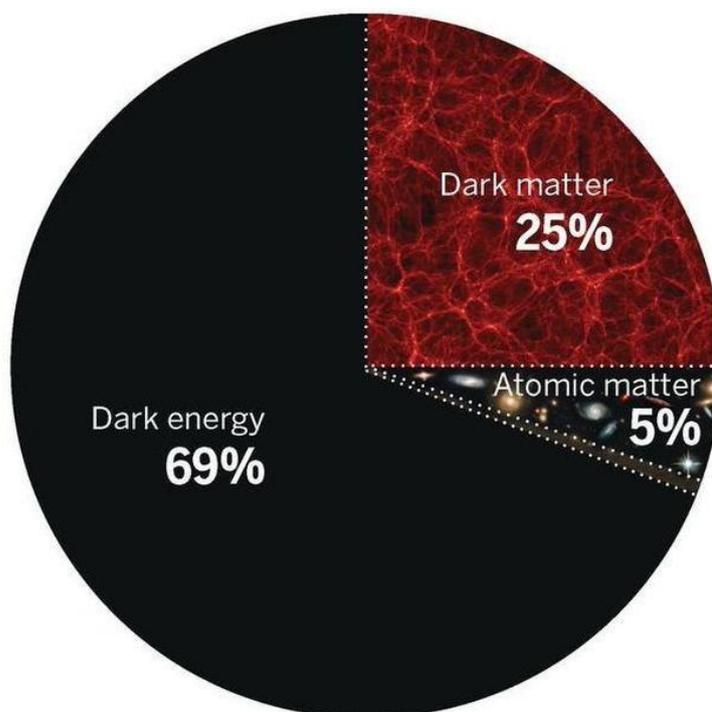


FIGURE 7.20 – Composition de l'Univers, à ce jour, en termes de densité d'énergie normalisées. Les pourcentages correspondent approximativement à la valeur donnée par la Cosmologie de Planck 2018 [1].

À l'heure actuelle, la meilleure description de l'Univers est celle du modèle appelé  $\Lambda$ CDM. Celui-ci modélise l'Univers comme plat ( $K = 0$ ), homogène, isotrope, et régi par la Relativité Générale.

À partir d'un état initial extrêmement dense et chaud (Big Bang), l'Univers s'étend et se refroidit du fait de cette expansion.  $\Lambda$ CDM est le modèle le plus simple pour expliquer nos observations actuelles. En particulier, il permet de faire des prédictions concernant le fond diffus cosmologique (CMB), l'abondance des éléments légers dans l'Univers, la présence de structures à grande échelle contenant des galaxies et l'expansion accélérée récente de l'Univers. Dans ce modèle, l'Univers est composé de la matière relativiste (photons), de la matière non relativiste (neutrinos, baryons et matière noire) et d'énergie sombre sous forme de constante cosmologique. Aujourd'hui, les densités d'énergie normalisées (noté  $\Omega$ ) de ces composantes sont montrées dans la figure 7.20. Dans ce bilan, le rayonnement contribue faiblement ( $\Omega_{\gamma,0} = 5.402 \times 10^{-5}$ ) mais ces poids ont beaucoup variés durant l'évolution de l'Univers. Ainsi, une période dominée par le rayonnement a laissé sa place à une période dominée par la matière puis d'expansion accélérée dominée par l'énergie sombre.

Une extension largement admise du modèle  $\Lambda$ CDM est la présence d'une phase d'inflation rapide de l'Univers à ces premiers instants. Elle est notamment responsable des perturbations initiales de la métrique de l'Univers qui ont, par leur évolution, formées les grandes structures de l'Univers sous forme de toile cosmique. L'inflation définit la pente ( $n_s$ ) et le niveau ( $A_s$ ) du spectre de puissance initial des perturbations de courbure.

Les 6 paramètres du modèle  $\Lambda$ CDM plat minimal sont les suivants :

$$\Theta_{\Lambda\text{CDM}} = \{\Omega_{\text{b},0}; \Omega_{\text{cdm},0}; H_0; \tau; A_s; n_s\},$$

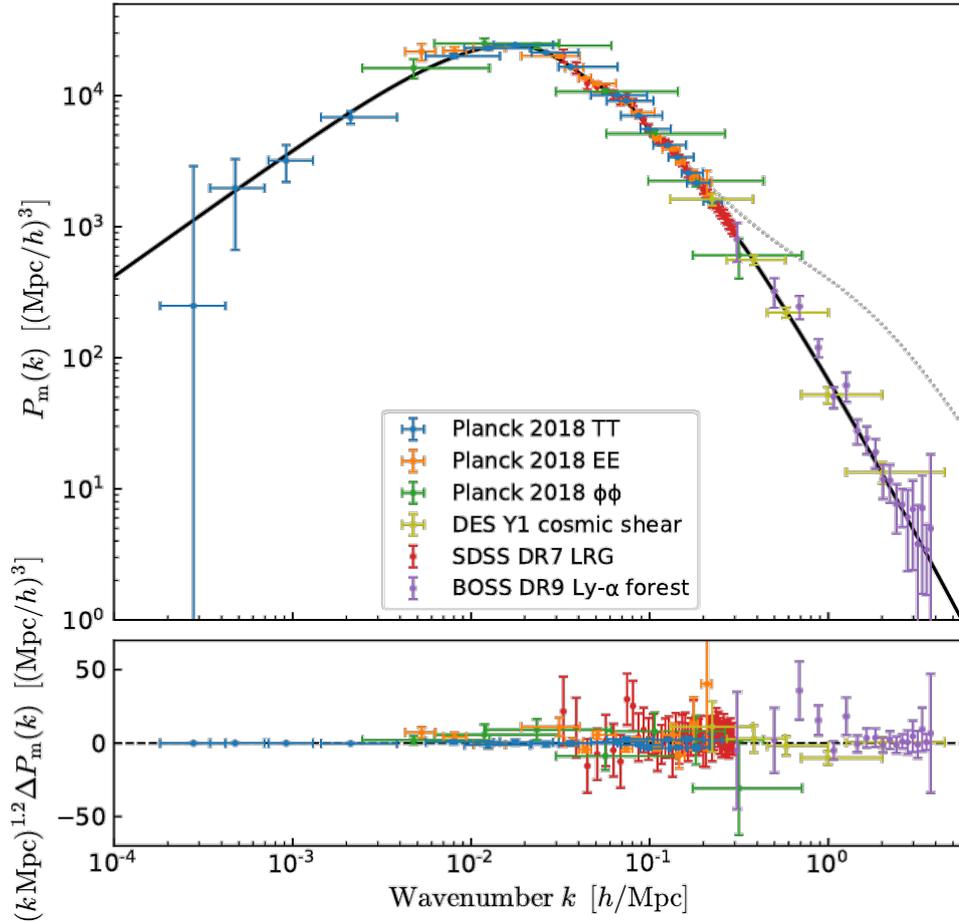


FIGURE 7.21 – Inférence du spectre de puissance linéaire de la matière au redshift  $z = 0$  à partir de différentes sondes cosmologiques, tirée de [2].

où  $\Omega_{b,0}$  et  $\Omega_{\text{cdm},0}$  sont les densités d'énergie normalisées des baryons et de la matière noire, et  $\tau$  est la profondeur optique du fond diffus cosmologique.

La prise en compte des perturbations de métrique nécessite de dériver les équations de la Relativité Générale dans le cadre de la théorie des perturbations. Lorsque nous considérons ces perturbations comme linéaires (en regardant les petites échelles ou les grands redshifts), cette théorie fournit le spectre de puissance linéaire de la matière en fonction du redshift. Ce spectre de puissance est calculé avec le code de Boltzmann CLASS [3, 4] et est représenté en ligne continue noire dans la figure 7.21.

Les points sur cette figure correspondent aux inférences réalisées par différentes sondes cosmologiques (fond diffus cosmologique, lentillage gravitationnel faible, clustering des galaxies et forêt Lyman- $\alpha$ ). La construction du modèle  $\Lambda$ CDM dépend avant tout de ces observations. Il existe de nombreuses autres observables cosmologiques. Les sondes énumérées ci-dessus sont très complémentaires. La combinaison de plusieurs types d'observations, est un outil puissant pour réduire l'impact des systématiques liées à chaque type de sondes, et pour améliorer les contraintes cosmologiques en exploitant leurs dégénérescences.

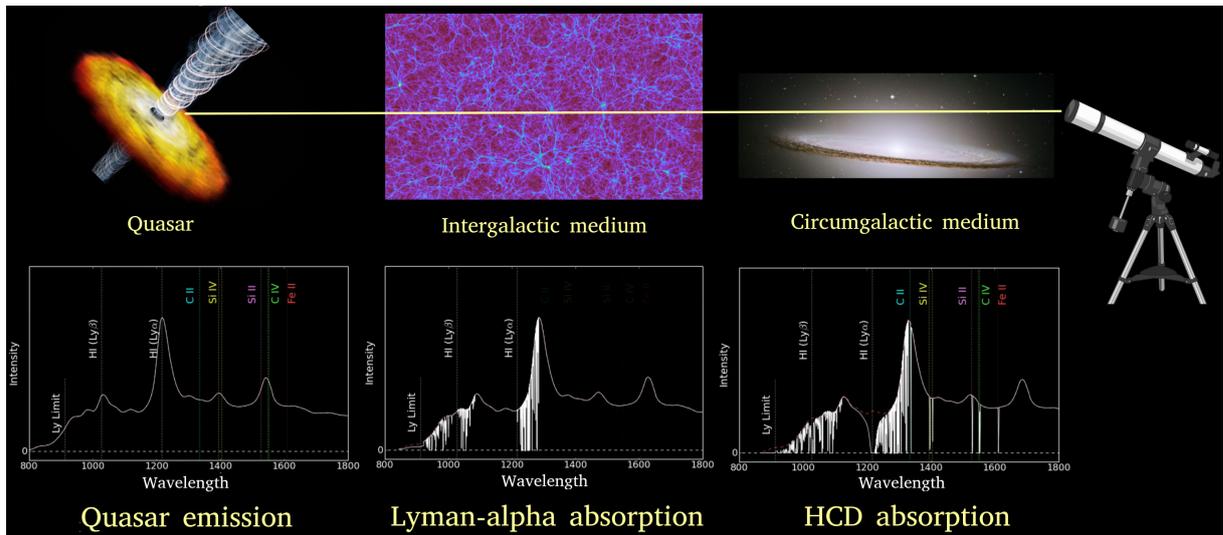


FIGURE 7.22 – Représentation schématique de la formation de la forêt Ly $\alpha$  adaptée de A. Pontzen. La lumière émise par le quasar est absorbée par l’hydrogène neutre du milieu intergalactique, formant ainsi la forêt Ly $\alpha$ .

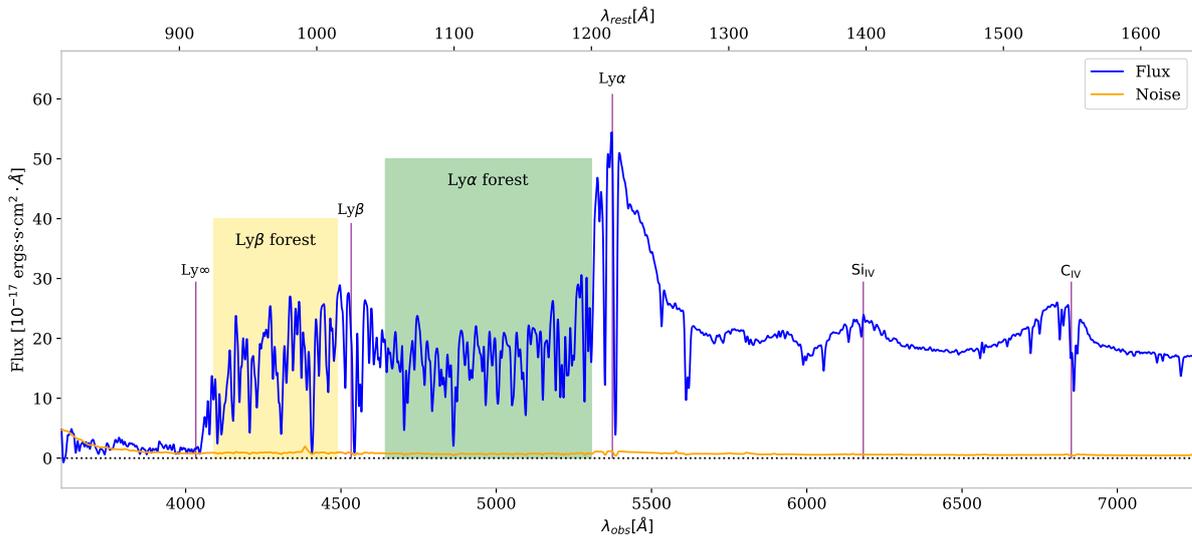


FIGURE 7.23 – Le spectre d’un quasar (RA = 217.263 °, DEC = -1.755 °) situé à un redshift  $z = 3.42$  mesuré par DESI. Le flux du quasar est représenté en bleu et son bruit instrumental en orange.

## II La forêt Lyman- $\alpha$

Cette thèse utilise la forêt Lyman- $\alpha$  (ci-après noté Ly $\alpha$ ) comme traceur cosmologique. Comparée aux traceurs ponctuels, tels que les galaxies, la forêt Ly $\alpha$  se distingue par sa capacité à sonder les fluctuations à petite échelle de la toile cosmique.

### Structure de la forêt Lyman- $\alpha$

La série de Lyman est l’ensemble des raies spectrales de l’hydrogène entre un niveau excité  $n \geq 2$  et son niveau fondamental. En particulier, la raie **Lyman- $\alpha$** , correspond à la transition

entre le premier état excité et l'état fondamental. Dans un cadre de repos, la longueur d'onde de cette transition est  $\lambda_\alpha = 1215,67 \text{ \AA}$ .

La forêt Ly $\alpha$  peut être observée en mesurant le spectre des quasars, qui sont parmi les objets astrophysiques les plus brillants connus à ce jour. La nature de ces objets est longtemps restée mystérieuse. Grâce aux progrès de l'astrophysique, on sait désormais que la lumière intense des quasars provient de la matière en rotation rapide autour d'un trou noir super-massif.

La figure 7.22 montre la construction schématique de la forêt Ly $\alpha$ . La lumière émise par un quasar traverse le milieu intergalactique, situé entre le quasar et l'observateur. Sur leur chemin, les photons sont partiellement absorbés par des éléments du milieu intergalactique. En particulier, les absorptions Ly $\alpha$  causées par l'hydrogène neutre, situé à différents redshifts, créent une séquence de lignes d'absorption sur le spectre du quasar, ce qui constitue la forêt Ly $\alpha$ . Un exemple de spectre de quasar avec une forêt Ly $\alpha$  visible est donné dans la figure 7.23.

La fraction de flux transmis, noté  $F$ , peut être exprimée en fonction de la profondeur optique de la forêt Ly $\alpha$  ( $\tau_\alpha$ ) suivant l'expression:

$$F(\nu) = \frac{f_{\text{out}}(\nu)}{f_{\text{in}}(\nu)} = e^{-\tau(\nu)}. \quad (7.55)$$

Pour calculer exactement cette profondeur optique, il faut connaître précisément le profil d'absorption de chaque absorbeur dans la toile cosmique et les vitesses particulières du milieu intergalactique. Une approximation pratique appelée "fluctuating Gunn-Peterson" (FGPA) [5, 6, 7, 8, 9] permet d'obtenir une expression approximée de cette profondeur optique aux grandes échelles ( $\gtrsim h^{-1} \cdot \text{Mpc}$ ):

$$\tau_\alpha \propto \left(\frac{1+z}{4}\right)^6 \left(\frac{\Omega_{b,0} h^2}{0.0125}\right)^2 \left(\frac{T_0}{10^4 \text{ K}}\right)^{-0.7} \left(\frac{\Gamma}{10^{-12} \text{ s}^{-1}}\right)^{-1} \left(\frac{H(z)}{100 \text{ km s}^{-1} \text{ Mpc}^{-1}}\right)^{-1} (1 + \delta_b)^\beta, \quad (7.56)$$

où  $\delta_b$  est le contraste de densité baryonique,  $\beta = 2 - 0,7(\gamma - 1)$ ,  $T_0$  est la température du milieu intergalactique à densité moyenne,  $\Gamma$  est le taux de photo ionisation, et  $h$  la constante de Hubble réduite. Aux grandes échelles, la profondeur optique de la forêt Ly $\alpha$  et les fluctuations baryoniques sous-jacentes sont donc reliées par une loi de puissance.

La forêt Ly $\alpha$  est contaminée par plusieurs effets astrophysiques, appelés systématiques. Les premiers sont les objets "high column density" (HCD) qui apparaissent lorsque la lumière des quasars traverse des régions à forte concentration d'hydrogène neutre. C'est le cas des régions à la périphérie des galaxies qui sont appelées milieu circum-galactique comme représenté dans la figure 7.22. De plus, il existe une forêt associée pour chaque autre élément absorbant (métal) présent dans le milieu intergalactique. Ces métaux contaminent la forêt Ly $\alpha$  lorsque leur transition est à une longueur d'onde plus élevée que la ligne d'émission Ly $\alpha$ . Enfin, certains quasars appelés "broad absorption line" (BAL) possèdent des absorptions larges décalées vers le bleu par rapport aux émissions des quasars. Ils peuvent également polluer les mesures utilisant la forêt Ly $\alpha$ .

### Statistiques utilisant la forêt Ly $\alpha$

Le but des mesures statistiques de la forêt Ly $\alpha$  est de calculer la fraction du contraste du flux transmis de la forêt Ly $\alpha$ , appelée **contraste Ly $\alpha$** . Cette observable peut être définie à partir du flux  $f$ , de la fraction de flux transmis  $F$  et de l'émission continue du quasar  $C_q$  suivant :

$$\delta_F(\lambda) = \frac{f(\lambda)}{C_q(\lambda, z_q)F(\lambda)} - 1. \quad (7.57)$$

À partir de la mesure du contraste Ly $\alpha$ , il est possible de définir plusieurs outils statistiques tels que la fonction de corrélation comme la fonction de corrélation croisée à deux points (également exprimée corrélation croisée) avec un autre traceur  $Y$  :

$$\xi_{\alpha Y}(\vec{r}) = \langle \delta_F(\vec{x})\delta_Y(\vec{x} + \vec{r}) \rangle_{\vec{x}}. \quad (7.58)$$

La forêt Ly $\alpha$  peut également être modélisée dans l'espace de Fourier. Le spectre de puissance tridimensionnel de la forêt Ly $\alpha$  ( $P_\alpha$ ) est défini par

$$P_\alpha(k)\delta_D^{(3)}(\vec{k} + \vec{k}') = \frac{1}{(2\pi)^3} \langle \delta_F(\vec{k})\delta_F(\vec{k}') \rangle. \quad (7.59)$$

La forêt Ly $\alpha$  est un traceur cosmologique à la géométrie particulière, car elle est fortement anisotrope. Par conséquent, il est intéressant de se concentrer sur l'étude des corrélations le long de la ligne de visée. C'est le cas du spectre de puissance unidimensionnel de la forêt Ly $\alpha$  ( $P_{1D,\alpha}$ ) qui peut être relié à  $P_\alpha$  :

$$P_{1D,\alpha}(k_{\parallel}) = \int \frac{d\vec{k}_{\perp}}{(2\pi)^2} P_\alpha(\vec{k}). \quad (7.60)$$

### III Modélisation numérique de la forêt Lyman- $\alpha$

Les simulations cosmologiques sont essentielles lorsque les modèles analytiques deviennent inefficaces. Par exemple, il est impossible de modéliser analytiquement les plus petites échelles de la forêt Ly $\alpha$ . De plus, elle nécessite de modéliser les propriétés de la matière baryonique qui font appel à une physique complexe.

#### III.1 Mocks pour les études des systématiques

Une première catégorie de simulation numérique est la création de données synthétiques, également appelées **mocks**. Leur but est de mimer une grande quantité de données, et l'évolution des fluides cosmologiques est généralement réalisée avec des méthodes approximatives. Les mocks sont très proches de l'analyse des données et sont conçus pour étudier l'impact des systématiques instrumentales, analytiques et théoriques. Dans cette thèse, j'ai utilisé deux types de mocks log-normaux pour caractériser l'impact des systématiques sur le spectre de puissance unidimensionnel, la tomographie et la corrélation croisée avec les vides.

Un mock log-normal  $\text{Ly}\alpha$  nécessite une méthode pour générer un champ de densité, et un modèle reliant ce champ à la forêt  $\text{Ly}\alpha$ . Dans notre cas, les mocks utilisent des champs aléatoires gaussiens (GRF) pour générer le champ de densité de la matière. La forêt  $\text{Ly}\alpha$  est ensuite construite de l'approximation **log-normale** [10, 11].

### SaclayMocks

J'ai généré des données synthétiques avec le logiciel **SaclayMocks**<sup>2</sup> afin d'interpréter les mesures des sections VI et VII. Ce logiciel modélise les champs cosmologiques comme des champs aléatoires gaussiens. Les **SaclayMocks** ont été développés pour reproduire les propriétés statistiques observées de la forêt  $\text{Ly}\alpha$  à des échelles relativement grandes [12, 13, 14, 15]. Ces mocks peuvent reproduire les champs observés des relevés eBOSS ou DESI.

Les champs gaussiens (densité, vitesse) sont générés dans une boîte de l'espace de Fourier. Les modes de cette boîte sont multipliés par plusieurs ensembles de poids appropriés de sorte que, après application d'une transformée de Fourier tridimensionnelle inverse, ils produisent plusieurs boîtes : une boîte de fluctuations de densité de la matière,  $\delta_L$ , à  $z = 0$  ; six boîtes de composantes de gradient de vitesse à  $z = 0$  ; et une boîte de fluctuations de densité des quasars à  $z = 2, 3$ . Un catalogue de quasars est obtenu en plaçant un quasar dans chaque cellule de cette dernière boîte avec une probabilité proportionnelle au logarithme de la densité. Afin de reproduire la dépendance en redshift des observations  $\text{Ly}\alpha$ , la relation FGPA (équation 7.56) est davantage paramétrée :

$$F = \exp \left[ -a_{\text{GP}}(z) \exp \left( b_{\text{GP}}(z) G(z) (\delta_L + \delta_S + c_{\text{GP}}(z) \eta_{\parallel}) \right) \right]. \quad (7.61)$$

Dans cette expression, le champ  $\delta_S$  est réglé pour que chaque forêt  $\text{Ly}\alpha$  suive le spectre de puissance unidimensionnel observé. Le gradient de vitesse le long de la ligne de visée,  $\eta_{\parallel}$ , est obtenu à partir des boîtes de composantes de gradient de vitesse. Les paramètres  $a(z)$ ,  $b(z)$ , et  $c(z)$  sont dimensionnés pour reproduire les paramètres  $\text{Ly}\alpha$  observés. Enfin,  $G(z)$  est le facteur de croissance linéaire.

À partir du catalogue de quasars, les lignes de visée sont tirées dans les différentes boîtes, et la fraction de flux transmis est calculée à l'aide de l'équation 7.61. Il est également possible d'obtenir le champ de matière à partir d'une transformation de  $\delta_L$ .

### DESI-Lite

Les besoins particuliers de la mesure du spectre de puissance unidimensionnel de la forêt  $\text{Ly}\alpha$  (section V) ont conduit à la création de mocks dédiés appelés **DESI-Lite** [16]. Dans ce cas, modéliser la distribution tridimensionnelle complète n'est pas nécessaire, les lignes de visée peuvent être considérées comme indépendantes les unes des autres. Nous imposons seulement que tous les contrastes  $\text{Ly}\alpha$  générés suivent un spectre de puissance donné, en accord avec la mesure de  $P_{1D,\alpha}$  dans [17]. Pour être plus précis, plusieurs transformations sont appliquées aux contrastes  $\text{Ly}\alpha$  pour obtenir la dépendance en redshift de  $P_{1D,\alpha}$ . En utilisant les calculs détaillés dans [16], il est possible d'obtenir précisément le spectre de puissance unidimensionnel que suivent toutes les forêts  $\text{Ly}\alpha$  générées.

2. <https://github.com/igmhub/SaclayMocks>

Les fichiers de fraction de flux transmis générés avec `SaclayMocks` ou `DESI-Lite` ont un format commun. Le logiciel `quickquasars` [18], transforme les fractions de flux transmis en spectres de quasar en modélisant les systématiques observationnelles et astrophysiques.

## III.2 Simulations cosmologiques

Les mocks sont inadéquats pour fournir des prédictions précises et quantifier l'impact des paramètres cosmologiques sur les observables. Dans ce cas, les **simulations cosmologiques** sont nécessaires. Ces dernières modélisent les composants cosmologiques comme des particules ou des fluides, et les font évoluer dans le temps en considérant leurs interactions.

Les simulations réalisées dans cette thèse utilisent le code hydrodynamique à grille `Nyx`<sup>3</sup> [21, 22]. Ce logiciel modélise les baryons comme un gaz idéal sur une grille eulérienne. La dynamique du gaz est résolue à l'aide d'une méthodologie de volumes finis du second ordre. La matière noire est modélisée par des particules dans un schéma d'interaction à N-corps. `Nyx` inclut également les processus physiques essentiels à la modélisation de la forêt Ly $\alpha$ .

### Simulations pour l'estimation de $P_{1D,\alpha}$

Pour calculer la dépendance de  $P_{1D,\alpha}$  par rapport aux paramètres cosmologiques, un **émulateur de processus gaussien** est utilisé. Ce type d'approche permet de couvrir avec une précision homogène l'espace des paramètres, même éloignés. Une grille de simulations est générée avec des paramètres disposés dans un échantillonnage en hypercube latin. Un processus gaussien est ensuite entraîné à émuler chacun des modes  $P_{1D,\alpha}(z, k)$  pour de nouveaux ensembles de paramètres.

La grille de simulation est constituée de 14 réalisations de taille 120 Mpc, avec 4096<sup>3</sup> éléments de résolution. Une représentation du champ de densité d'une simulation est donnée dans la figure 7.24. Les paramètres cosmologiques des différentes réalisations correspondent aux points de l'échantillonnage en hypercube latin représentés dans la figure 7.25. Les paramètres cosmologiques s'étendent sur une plage de  $\pm 5 \sigma$  autour des valeurs de Planck [1]. L'exploitation de cette grille pour les contraintes cosmologiques est en cours. Elle sera employée pour l'interprétation de la mesure de  $P_{1D,\alpha}$  avec les données de DESI (section V).

### Prédiction de $P_\alpha$ par simulations

Le spectre de puissance unidimensionnel est la seule mesure réalisée avec les données existantes, dans l'espace de Fourier, de la forêt Ly $\alpha$ . Avec l'avènement de DESI [23, 24] et WEAVE-QSO [25], la mesure du spectre de puissance tridimensionnel de la forêt Ly $\alpha$  ( $P_\alpha$ ) pourrait être réalisable. En effet, en raison de la forme très anisotrope de la forêt Ly $\alpha$ , cette mesure n'est possible que pour une grande densité de lignes de visée qui fournit une résolution spatiale similaire dans les directions transversale et radiale.  $P_\alpha$  a déjà été calculé sur des simulations hydrodynamiques dans [20, 26, 27]. Cependant, les boîtes de simulation utilisées dans ces études restent de taille limitée.

Pour calculer  $P_\alpha$  jusqu'à de très grandes échelles, j'ai codirigé un projet en collaboration avec des chercheurs du DPhP, du DAp, et du LBNL dans le but de produire des simulations

3. <https://amrex-astro.github.io/Nyx>

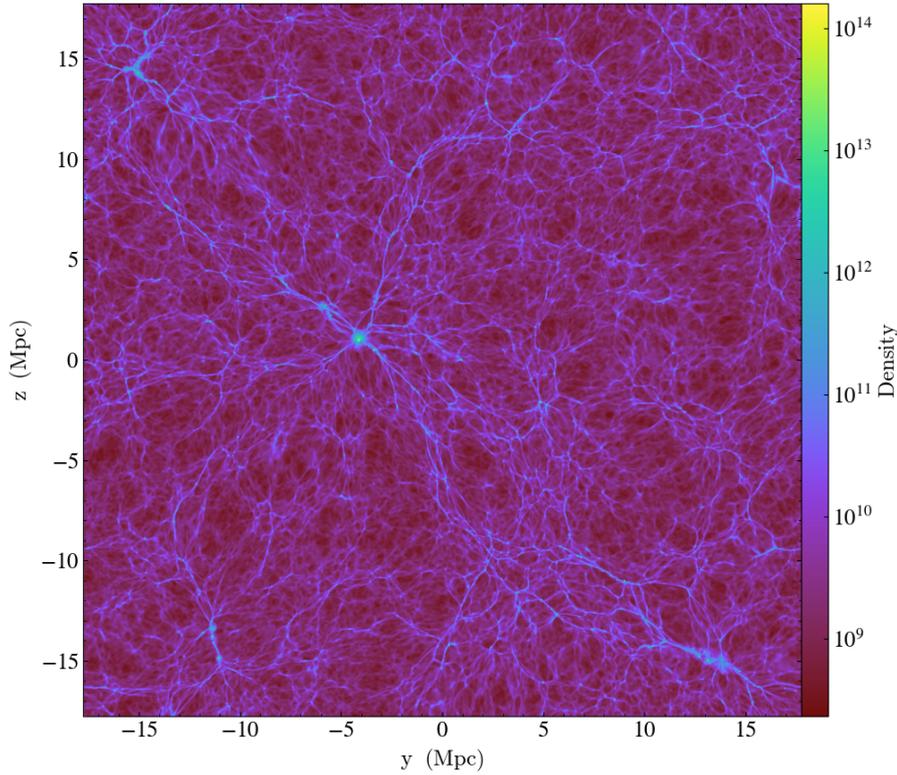


FIGURE 7.24 – Tracé de la densité de matière d’une simulation de la grille à  $z = 2,6$ . La taille en coordonnées comobiles est de 120 Mpc pour chaque direction. La densité de matière est exprimée en unité de  $M_{\odot} \cdot \text{Mpc}^{-3}$ .

hydrodynamiques de grand volume (jusqu’à  $500 h^{-1} \cdot \text{Mpc}$ ). Ces nouvelles simulations constituent un laboratoire idéal pour tester les modèles analytiques avec une précision suffisante tout en limitant les effets de la variance cosmique.

Nous avons utilisé la nouvelle implémentation GPU de Nyx et avons montré qu’elle était jusqu’à environ 100 fois plus rapide en temps d’exécution, compte tenu du nombre de GPU et de CPU utilisés. Nous avons réalisé un ensemble de simulations en faisant varier la taille et la résolution des réalisations pour tester l’impact de ces paramètres sur la mesure de  $P_{\alpha}$ . Un exemple de mesure est donné dans la figure 7.26 (gauche). Nous paramétrons  $P_{\alpha}$  en suivant [20, 26], tel que :

$$P_{\text{modèle},\alpha}(k, \mu) = b_{\alpha}^2 \left(1 + \beta_{\alpha} \mu^2\right)^2 P_{\text{m}}(k) D(k, \mu), \quad (7.62)$$

où  $b_{\alpha}$  est le biais  $\text{Ly}\alpha$ ,  $\beta_{\alpha}$  le paramètre de "redshift space distortions" (RSD) à grande échelle de la forêt  $\text{Ly}\alpha$ , et  $D(k, \mu)$  est une contribution non-linéaire modélisée suivant [20]. L’ajustement de ce modèle aux prédictions de  $P_{\alpha}$  dans les simulations fournit une mesure de  $b_{\alpha}$  et  $\beta_{\alpha}$ , montrée dans la figure 7.26 (droite).

Cette étude préliminaire fournit des résultats très encourageants. Dans un futur proche, notre objectif est d’améliorer la mesure présentée dans cette section afin d’obtenir une prédiction robuste des paramètres  $\text{Ly}\alpha$  linéaires et non linéaires.

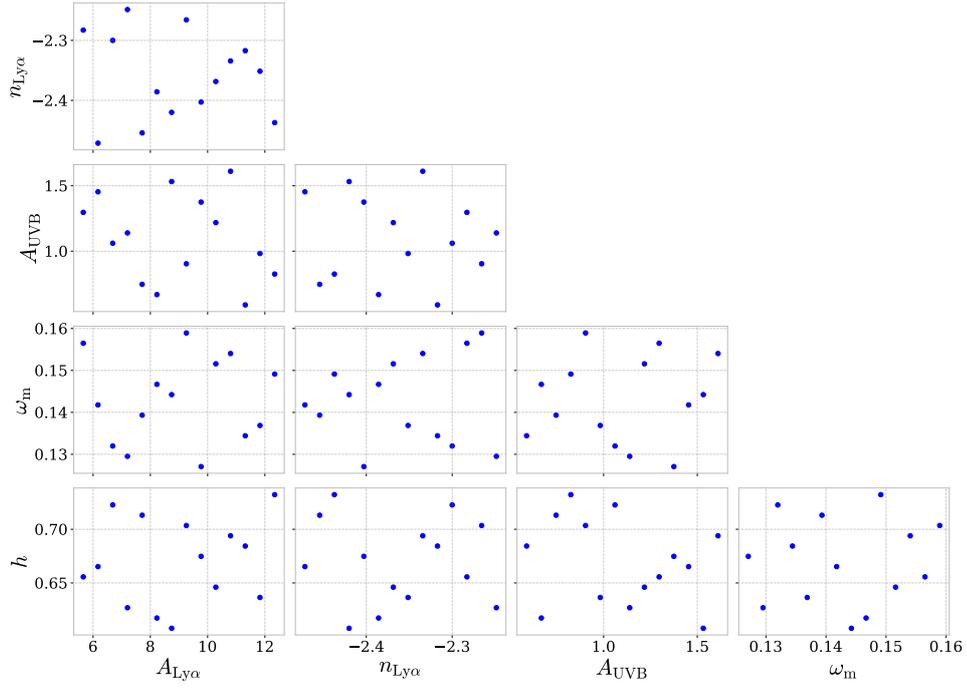


FIGURE 7.25 – Les points générés par l'échantillonnage en hypercube latin  $\Theta_{\text{sim}}$  qui correspondent aux paramètres cosmologiques et de fond UV des quatorze simulations [19].

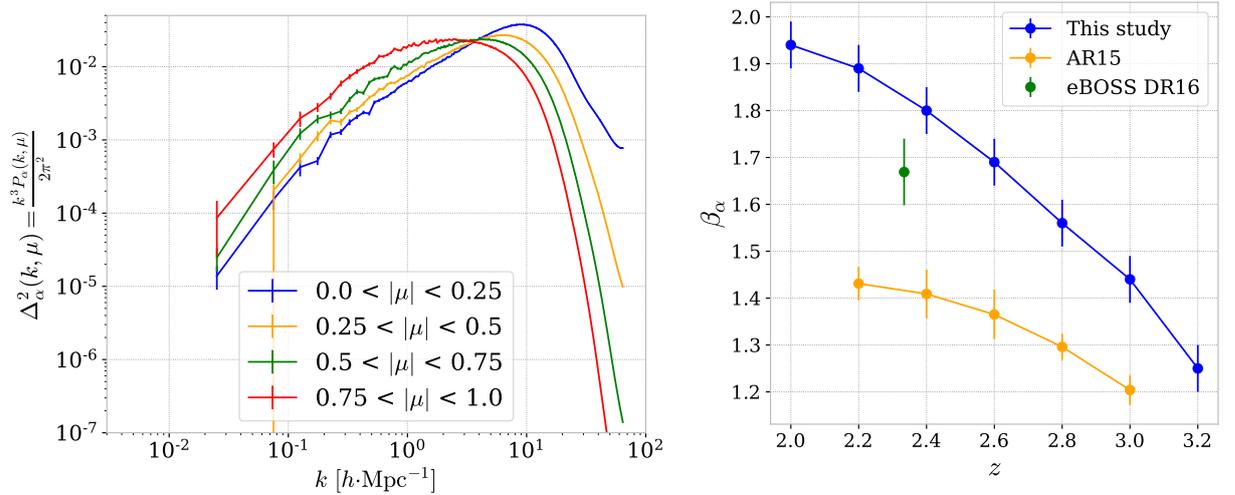


FIGURE 7.26 – (gauche) Spectre de puissance tridimensionnel sans dimension d'une simulation de  $125 h^{-1} \cdot \text{Mpc}$ , à  $z = 2.0$ . (droite) Valeurs ajustées en bleu pour les paramètres  $b_\alpha$  et  $\beta_\alpha$  sur le spectre de puissance tridimensionnel, en fonction du redshift. Les points jaunes (AR15) correspondent à la prédiction de [20] avec le même modèle d'ajustement. Le point vert est la mesure de ces paramètres avec les données eBOSS DR16 [15].

## IV Spectrographes multi-objets

Durant ma thèse, les spectres de quasars que j'ai utilisés proviennent d'instruments appelés **spectrographes multi-objets** qui sont composés des éléments suivants :

- **Télescope**: Le système optique du télescope concentre le flux lumineux d'un certain champ de vision dans le ciel.

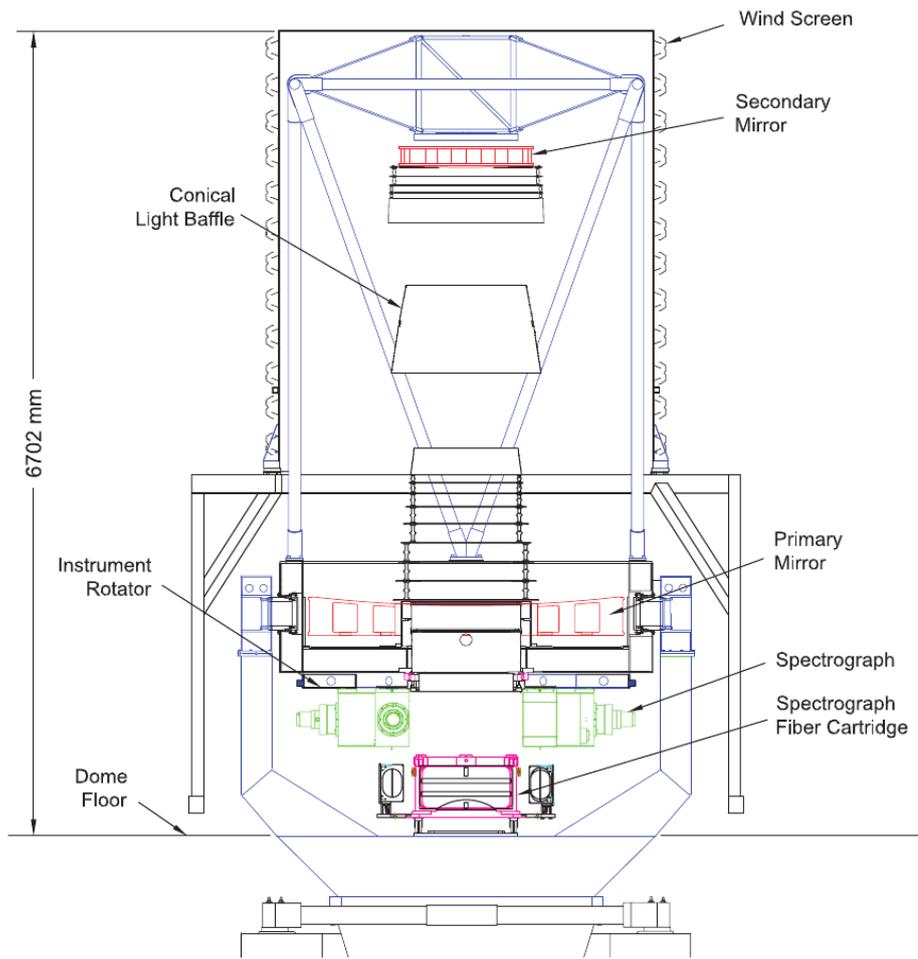


FIGURE 7.27 – Représentation schématique des différents composants du télescope SDSS issue de [28].

- **Fibres optiques:** Les fibres sélectionnent une cible sur le plan focal et dirigent le flux vers un spectrographe.
- **Spectrographe:** Il décompose la lumière des fibres en plusieurs longueurs d'onde grâce à un système optique dispersif. Cette lumière dispersée est ensuite capturée par des caméras.

### eBOSS

Le Sloan Digital Sky Survey (SDSS) a été imaginé dans les années 1980. Ce télescope est situé à l'observatoire d'Apache Point au Nouveau-Mexique. Le système optique du télescope a été utilisé pour effectuer des relevés photométriques et spectroscopiques, en commutant au point focal une caméra d'imagerie ou une plaque de fibre connectée à deux spectrographes. Le même télescope a donc été utilisé pour la sélection des cibles et la mesure des spectres correspondants [28, 30, 31]. Une vue schématique en coupe du télescope SDSS est donnée dans la figure 7.27. Ce télescope possède un miroir primaire de 2,5 mètres de diamètre et il a un champ de vision de  $3 \text{ deg}^2$ .

Le projet SDSS s'est déroulé en quatre phases achevées et est actuellement dans sa cinquième phase. Chaque phase dure environ 5 ans. Le **extended Baryon Oscillation Spectroscopic Survey (eBOSS)** [32] est le relevé SDSS utilisé dans cette thèse (quatrième phase). Il a permis d'augmenter les statistiques de chaque traceur (galaxies et quasars) par rapport aux observa-

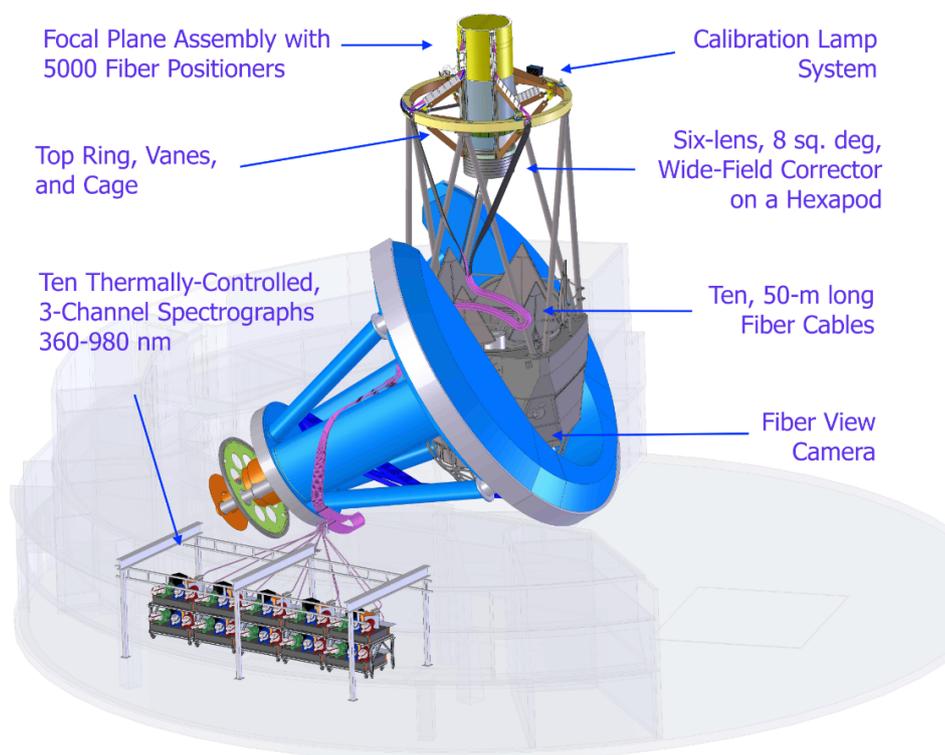


FIGURE 7.28 – Vue tridimensionnelle de l'instrument DESI monté à l'intérieur du télescope Mayall, issue de [29].

tions précédentes de SDSS. Les observations eBOSS ont conduit à des contraintes fortes sur le modèle  $\Lambda$ CDM et sur ses extensions possibles (neutrinos, modèles d'énergie sombre, gravité modifiée) [33]. Les forêts Ly $\alpha$  mesurées par les relevés BOSS et eBOSS sont utilisées dans les sections VI et VII.

## DESI

Le **Dark Energy Spectroscopic Instrument (DESI)** est un projet qui s'inscrit dans la continuité d'eBOSS. Il a démarré en octobre 2019. En raison de la pandémie de Covid-19, le télescope a connu un arrêt entre mars et octobre 2020. Après une phase de mise en service, le télescope a observé en continu à partir de décembre 2020, à l'exception d'un arrêt prévu à l'été 2021. L'instrument DESI est monté dans le télescope Mayall localisé en Arizona.

Le projet DESI est plus ambitieux que son prédécesseur et repose sur un changement de tous les matériels et logiciels utilisés précédemment. Une illustration globale de l'instrument et du télescope est présentée dans la figure 7.28. Le système global DESI (instrument et télescope) est composé de :

- **Télescope Mayall:** La structure métallique globale et le système optique primaire (miroir de 4 mètres).
- **Correcteur de mise au point primaire:** Le correcteur optique situé au sommet du télescope.
- **Système plan focal:** Système positionné sur le plan focal et qui est composé de 5000 fibres contrôlées par des robots.
- **Spectrographes:** Un module séparé contenant 10 spectrographes.

L'objectif de la collaboration DESI est d'observer les spectres de 35 millions d'objets extragalactiques sur un champ de  $14000 \text{ deg}^2$  en 5 ans. Ce projet vise à poursuivre les efforts de cartographie cosmique réalisés par SDSS, tout en augmentant drastiquement son pouvoir de contrainte sur le modèle  $\Lambda\text{CDM}$  et ses extensions possibles.

Une partie importante de cette thèse a été consacrée à la compréhension et à l'analyse des premières données de DESI. En particulier, ces données sont utilisées dans les sections V et VI.

## V Le spectre de puissance unidimensionnel de la forêt Lyman- $\alpha$

La distribution à petite échelle de la matière est encodée dans les fluctuations de la forêt Ly $\alpha$  le long de la ligne de visée. Le spectre de puissance unidimensionnel de la forêt Ly $\alpha$  ( $P_{1D,\alpha}$ ) est un outil puissant pour sonder cette distribution de la matière ainsi que l'état thermique du milieu intergalactique.

Mesurer  $P_{1D,\alpha}$  nécessite d'abord d'extraire le contraste Ly $\alpha$  après normalisation par le continuum des quasars. J'ai obtenu une mesure de  $P_{1D,\alpha}$  à partir de ces contrastes en utilisant la transformée de Fourier rapide. Cette méthode a été utilisée dans les analyses BOSS et eBOSS [17, 34] et l'estimateur correspondant est:

$$P_{1D,\alpha}(k) = \left\langle [P_{\text{raw}}(k) - P_{\text{noise}}(k)] \cdot \mathbf{R}^{-2}(k) \cdot \text{sinc}^2 \left( \frac{k\Delta\lambda_{\text{pix}}}{2} \right) \right\rangle - P_{\text{metals}}(k), \quad (7.63)$$

où  $P_{\text{raw}}$ ,  $P_{\text{metals}}$ , et  $P_{\text{noise}}$  sont respectivement les spectres de puissance associés aux absorptions dues à la forêt Ly $\alpha$ , au bruit et aux autres composants du milieu intergalactique (métaux). Le terme  $\mathbf{R}$  contient l'information de résolution des spectrographes de DESI et  $\lambda_{\text{pix}}$  est la séparation entre deux pixels d'un spectre.

Notre objectif est de calculer  $P_{1D,\alpha}$  à partir des premières données DESI, en suivant la même méthodologie que pour la dernière mesure eBOSS [34].  $P_{1D,\alpha}$  est une observable sensible aux propriétés instrumentales telles que le bruit et la résolution spectrale. Dans le cadre de ma thèse, j'ai caractérisé l'impact de l'estimation du bruit, du spectre de puissance des métaux, et des lignes d'émission de l'atmosphère sur la mesure de  $P_{1D,\alpha}$ . J'en ai déduit les corrections adéquates appliquées à l'estimateur donné par l'équation 7.63.

Pour tester les différentes étapes de la mesure de  $P_{1D,\alpha}$ , j'ai utilisé des mocks DESI-Lite spécialement conçus pour  $P_{1D,\alpha}$ , comme décrit dans la section III. Ces mocks sont utilisés pour caractériser l'impact du masquage des pixels (dû aux objets de type HCD ou aux émissions atmosphériques) et des effets résiduels tels que l'impact de l'ajustement du continuum sur le  $P_{1D,\alpha}$ . Ceci m'a permis de calculer des corrections empiriques à ces effets que j'ai ajouté à l'équation 7.63.

En considérant toutes les corrections définies précédemment, la mesure de  $P_{1D,\alpha}$  sur les premières données de DESI est montrée dans la figure 7.29. Cette mesure est effectuée pour 9 tranches de redshift allant de 2,2 à 3,8, et en utilisant un total de 21050 de sous-forêts (forêts Ly $\alpha$  découpées en trois). J'ai également réalisé une étude des incertitudes statistiques et systématiques associées à cette mesure.

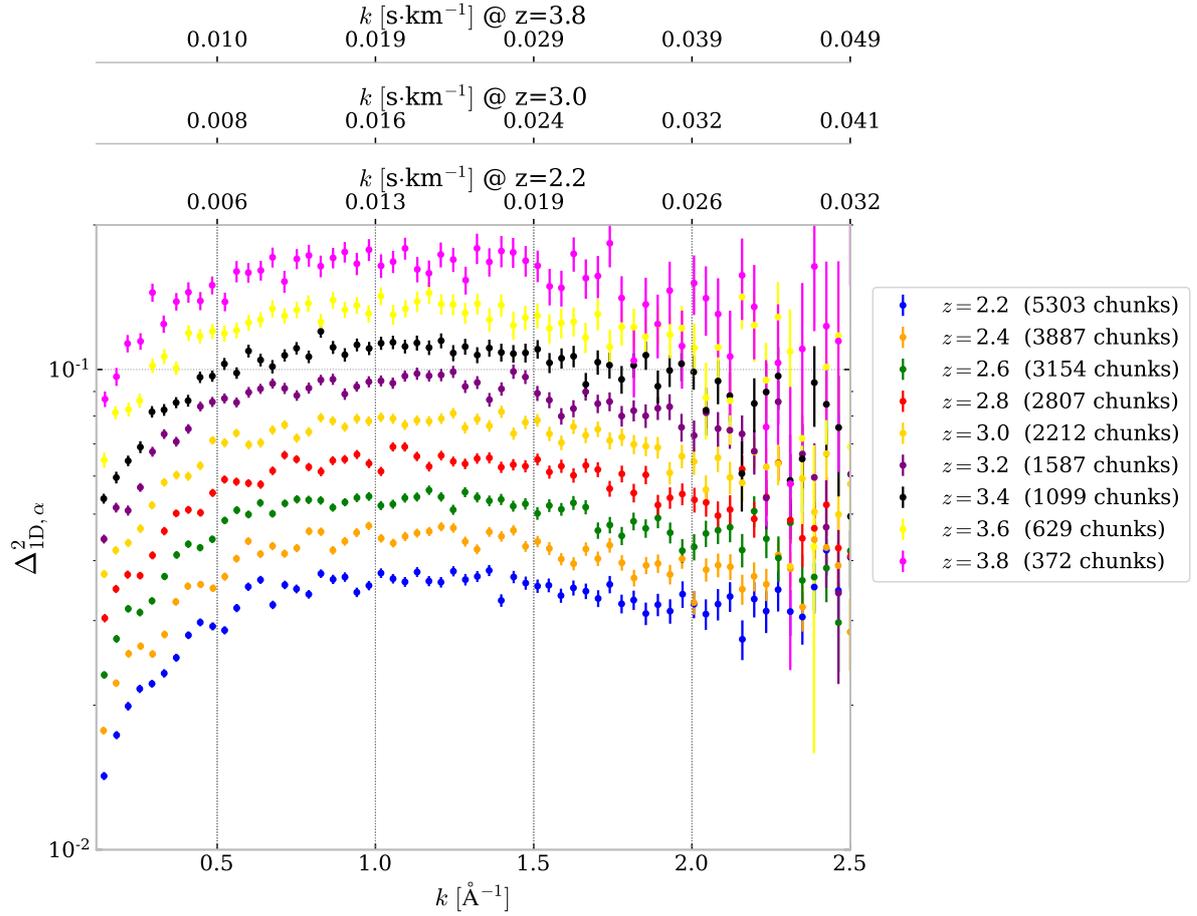


FIGURE 7.29 – Spectre de puissance unidimensionnel normalisé de la forêt Ly $\alpha$  ( $\Delta_{1D,\alpha}(k)$ ) utilisant les premières données de DESI, pour des redshift de  $z = 2, 2$  à  $z = 3, 8$ . Les barres d'erreur sont l'addition en quadrature des incertitudes systématiques et statistiques.

Ce travail démontre déjà des améliorations significatives par rapport aux résultats d'eBOSS. La précision statistique de cette première mesure ainsi que la gamme de nombre d'ondes couverte présente déjà un intérêt pour la Cosmologie. Dans un futur proche, nous prévoyons d'utiliser la grille de simulations de  $P_{1D,\alpha}$  présentée dans la section III pour obtenir des contraintes cosmologiques.

## VI Tomographie Lyman- $\alpha$

J'ai réalisé une interpolation entre les forêts Ly $\alpha$  pour créer une carte tridimensionnelle de l'absorption Ly $\alpha$ . Cette méthode est appelée tomographie Ly $\alpha$  [35]. J'ai utilisé les forêts Ly $\alpha$  d'eBOSS et je me suis concentré, en particulier, sur la partie la plus dense et la plus homogène de ce relevé : une bande étroite appelée Stripe 82, d'une superficie de 220 deg $^2$ , située dans le plan équatorial. J'ai utilisé les spectres de 8200 quasars, ce qui correspond à une densité de surface de 37 quasars par deg $^2$ .

Pour construire une carte tomographique, j'ai appliqué un filtre de Wiener tel qu'implémenté par CLAMATO [36, 37]. Cet algorithme effectue une interpolation avec un noyau gaussien en tenant compte du bruit des forêts Ly $\alpha$ . La figure 7.30 représente une tranche de la carte

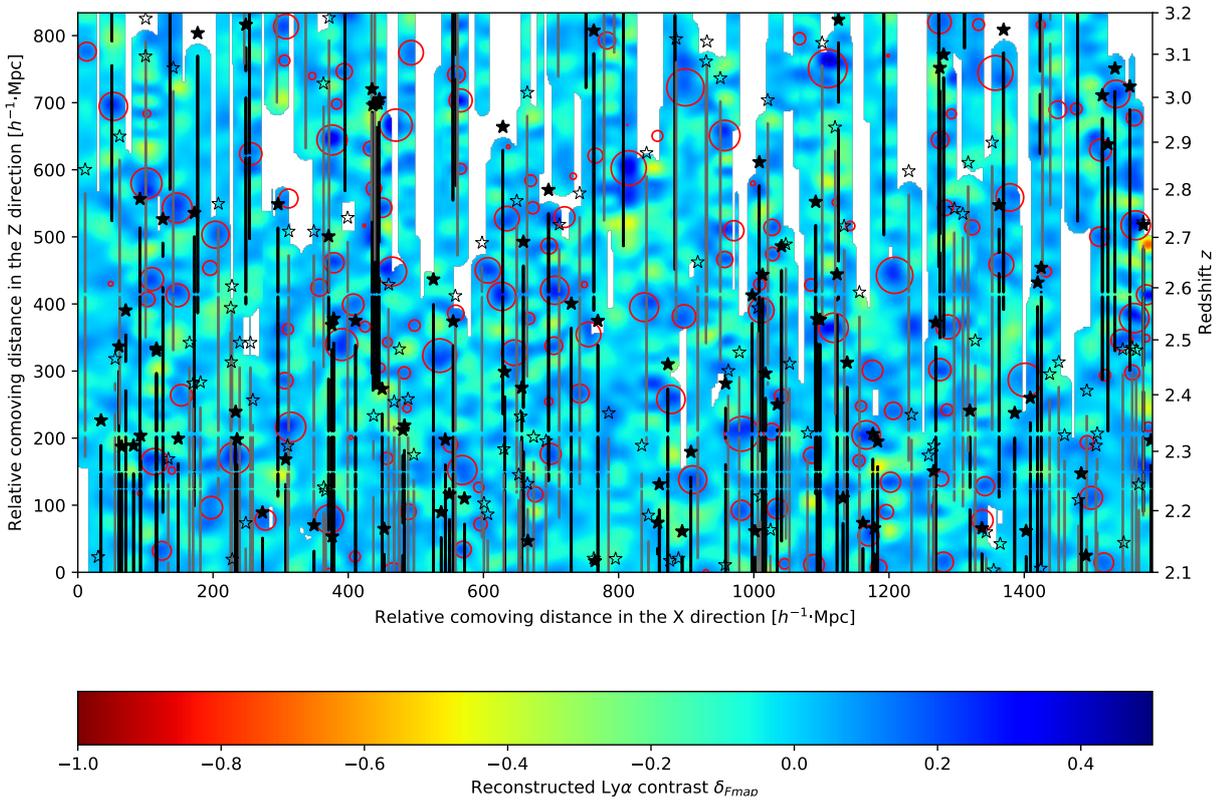


FIGURE 7.30 – Tranche de la carte tomographique de contraste Ly $\alpha$  calculée à partir des données eBOSS observées dans le Stripe 82. Les lignes de visée sont représentées par des traits noirs, les quasars par des étoiles et les vides par des cercles rouges. Tous les quasars et les lignes de visée représentés sont situés dans une tranche de  $20 h^{-1}\cdot\text{Mpc}$  d'épaisseur.

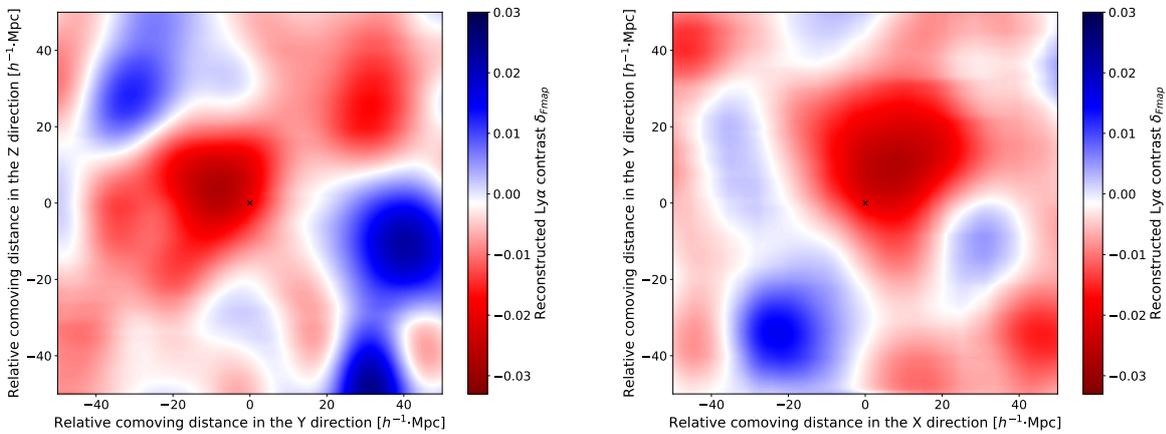


FIGURE 7.31 – Carte tomographique Ly $\alpha$  de DESI empliée sur les positions des GLBs. (gauche) Vue dans le plan  $Y - Z$ . (droite) Vue dans le plan  $X - Y$ .

tomographique calculée à partir des données fournies par eBOSS [38]. Le principal paramètre de l'algorithme tomographique est la longueur de corrélation du noyau gaussien que j'ai choisi égale à  $13 h^{-1}\cdot\text{Mpc}$ , ce qui correspond à la séparation moyenne entre les lignes de visée. Cette reconstruction tomographique constitue la première carte tridimensionnelle à grand volume et à haut redshift des fluctuations de densité de la matière.

En parallèle, j'ai utilisé les `SaclayMocks` décrit dans la section III pour tester les algorithmes de reconstruction tomographique. En extrayant le champ de densité associé à la carte tomographique reconstruite, j'ai quantifié dans quelle mesure cette carte retrace la densité de matière. La corrélation entre ces deux champs est de 34%, ce qui est en accord avec les précédentes publications [39].

La carte tomographique obtenue à partir des données eBOSS peut être utilisée pour plusieurs applications. Tout d'abord, un empilement de la carte tomographique autour des quasars révèle un signal clair centré sur la position du quasar : il s'agit d'une représentation alternative de la corrélation croisée entre les quasars et la forêt Ly $\alpha$  telle qu'étudiée dans [15]. Ensuite, j'ai identifié huit candidats proto amas en sélectionnant les sur-densités dans la carte, et en exigeant qu'ils soient traversés par un grand nombre de lignes de visée. Enfin, l'application sur laquelle je me suis concentré est la recherche de vides. J'ai développé un algorithme de recherche de vides sphériques, et j'ai créé le plus grand catalogue de grands vides à haut redshift. Pour construire ce catalogue, j'ai sélectionné uniquement les vides traversés par plusieurs lignes de visée, et dont le rayon est supérieur à  $7 h^{-1} \cdot \text{Mpc}$ .

Dans un second temps, j'ai appliqué les algorithmes de reconstruction tomographique aux premières données DESI, sur un champ particulier. Dans ce champ, des observations de galaxies Lyman-break (GLBs) ont été réalisées dans le cadre d'une campagne dédiée aux cibles secondaires, dans laquelle j'ai été impliqué. Les GLBs sont des galaxies localisées à hauts redshifts, qui forment beaucoup d'étoiles. Leurs spectres sont presque entièrement absorbés pour les longueurs d'onde inférieures à la limite de Lyman ( $\lambda_\infty = 911.75 \text{ \AA}$ ). Les GLBs sont un candidat majeur pour un futur relevé successeur de DESI.

J'ai empilé les positions des GLBs observées sur la carte tomographique obtenue avec les quasars de DESI. Le résultat en utilisant 288 GLBs est montré dans la figure 7.31. Il existe un excès de matière, tel que tracé par la carte tomographique Ly $\alpha$ , à la position de la GLB. La présence d'un signal avec une statistique aussi faible est extrêmement encourageante.

## VII Corrélation croisée entre les vides cosmiques et la forêt Lyman- $\alpha$

J'ai utilisé le catalogue de vides présenté dans la section précédente pour étendre les études de corrélation galaxie-vide [40, 41] à des redshifts  $z > 2$ . Pour ce faire, j'ai mesuré la corrélation croisée entre les centres des vides et les forêts Ly $\alpha$  utilisés pour la tomographie [42]. L'estimateur que j'ai utilisé pour cette mesure est :

$$\xi_{v\alpha}(A) = \frac{\sum_{(i,j) \in A} w_i \delta_{F,i}}{\sum_{(i,j) \in A} w_i}, \quad (7.64)$$

où  $j$  correspond à l'indice du vide et  $i$  à l'indice du pixel de la forêt Ly $\alpha$ . La séparation entre deux paires est caractérisée par une longueur  $r$  et l'angle cosinus  $\mu$ . Les poids  $w_i$  des pixels associés  $\delta_{F,i}$  dépendent du bruit et du redshift.

Cette étude vise à utiliser la forme angulaire de  $\xi_{v\alpha}$  pour observer l'effet des "redshift space distortions" (RSD) autour des vides aux redshifts  $z > 2$ . La fonction  $\xi_{v\alpha}$  telle qu'elle est calculée

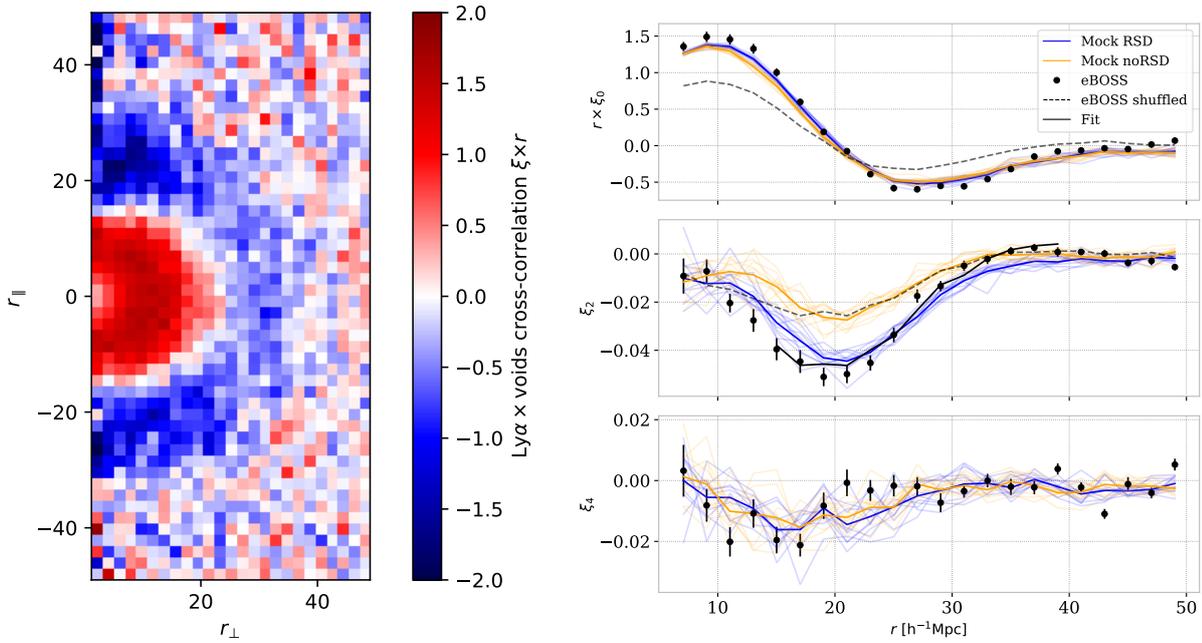


FIGURE 7.32 – (gauche) Représentation 2D de la corrélation croisée Ly $\alpha$ -vide dans les coordonnées  $A = (r_{\parallel}, r_{\perp})$  des données d’eBOSS sur le Stripe 82. (droite) Monopoles, quadrupoles et hexadecapoles de la corrélation croisée Ly $\alpha$ -vide pour les données d’eBOSS (points noirs), les mocks avec RSD (bleu) et sans RSD (orange). Les courbes noires en pointillés montrent le monopole et le quadrupole moyens mesurés après le remaniement des données eBOSS. La courbe noire continue montre l’ajustement du quadrupole d’eBOSS avec l’équation 7.65.

à partir des données de Stripe 82 est illustrée dans la figure 7.32 (gauche), et une décomposition en multipôles sur la base des polynômes de Legendre,  $\xi_{v\alpha,\ell}(r)$  pour  $\ell = 0, 2, 4$ , est présentée sur la figure 7.32 (droite). Pour interpréter la mesure, j’ai également appliqué cette méthode sur un ensemble de *SaclayMocks* présentant la même géométrie que le Stripe 82. Tout d’abord, j’ai utilisé une série de mocks pour évaluer l’impact de certaines systématiques instrumentales et astrophysiques sur la mesure (ajustement du continuum des quasars, métaux, HCD, bruit). La principale conclusion est que pour cet ensemble de données, l’impact des systématiques considérées est faible par rapport aux fluctuations statistiques.

Ensuite, j’ai calculé les multipôles de  $\xi_{v\alpha}$  pour un ensemble de onze mocks, ainsi que pour un ensemble supplémentaire de mocks compagnons, étiquetés "noRSD", pour lesquels les RSD ne sont pas pris en compte dans le calcul de l’absorption Ly $\alpha$ . Les résultats sont présentés dans la figure 7.32 (droite). Les données démontrent clairement le signal RSD, avec une signification statistique de  $10, \sigma$ .

Pour interpréter ce résultat, j’ai utilisé un modèle de vide simple. Il est basé sur le profil de vitesse moyen autour des vides dans le régime linéaire. Dans ce modèle, le monopole et le quadrupole de la corrélation croisée Ly $\alpha$ -vide sont reliés par une relation impliquant le paramètre RSD du vide  $\beta$ . Cette relation est très similaire au cas de la corrélation galaxie-void [41] :

$$\xi_{v\alpha,2}(r) = \left( \frac{2\beta}{3 + \beta} \right) \left( \xi_{v\alpha,0}(r) - \bar{\xi}_{v\alpha,0}(r) \right). \quad (7.65)$$

Comme on peut le voir sur la figure 7.32 (droite), la corrélation croisée présente un quadrupole non nul même en l'absence de RSD. Cette caractéristique provient de la géométrie particulière des données Ly $\alpha$ , qui affecte les positions des vides reconstruits. Le contraste moyen du flux de la carte tomographique construite avec le filtre de Wiener est plus faible aux endroits les plus éloignés des lignes de visée. Cela déplace les positions reconstruites des centres des vides plus près de la ligne de visée. J'ai inclus cet effet de manière simple dans le modèle donné par l'équation 7.65. J'ai également remanié les données pour confirmer l'origine géométrique de cet effet.

En ajustant ce modèle corrigé à la corrélation croisée Ly $\alpha$ -vide sur les données, j'ai obtenu un paramètre RSD  $\beta = 0.52 \pm 0.05$ . Une étude complète du biais de vitesse à partir de simulations hydrodynamiques est probablement nécessaire pour interpréter complètement cette mesure.

## Conclusion

Le modèle  $\Lambda$ CDM présenté dans la section I nous offre actuellement une description globalement unifiée de l'Univers. Dans ma thèse, j'ai essayé de contribuer à cette compréhension de l'Univers en étudiant la forêt Ly $\alpha$ , détaillée dans la section II. À cette fin, j'ai travaillé sur les données fournies par les télescopes SDSS et DESI présentées dans la section IV.

Dans la section III, j'ai présenté les outils de simulation utiles à l'analyse de la forêt Ly $\alpha$ , et en particulier, les mocks log-normaux et les simulations hydrodynamiques auquel j'ai contribué. Une partie importante de cette thèse est consacrée à la mesure du spectre de puissance unidimensionnel ( $P_{1D,\alpha}$ ) à partir des premières données de DESI, présentée dans la section V. À partir des spectres de quasars d'eBOSS mesurés dans le champ spécifique Stripe 82, j'ai construit, dans la section VI, une carte tomographique Ly $\alpha$  [38]. Dans la section VII, j'ai utilisé cette carte pour réaliser une mesure de la corrélation croisée Ly $\alpha$ -vide à un redshift médian  $z = 2.49$  [42].

La plupart des méthodes utilisées dans cette thèse peuvent encore être améliorées. De plus, des études supplémentaires sur les simulations hydrodynamiques pourraient potentiellement maximiser l'apport scientifique de ces travaux. Finalement, Les données des relevés DESI [23, 24] et WEAVE-QSO [25] permettront d'améliorer toutes les mesures développées dans cette thèse. En effet, l'augmentation de la densité et du nombre de forêts Ly $\alpha$  observées permettra une mesure plus précise de  $P_{1D,\alpha}$ , une carte tomographique mieux résolue et plus grande, et une augmentation significative de la statistique de vide.

## Bibliography

- [1] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astronomy & Astrophysics* **641** (2020) A6.
- [2] S. Chabanier, M. Millea and N. Palanque-Delabrouille, *Matter power spectrum: from Ly $\alpha$  forest to CMB scales*, *Monthly Notices of the Royal Astronomical Society* **489** (2019) 2247.
- [3] J. Lesgourgues, *The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview*, *arXiv:1104.2932 [astro-ph]* (2011) .
- [4] D. Blas, J. Lesgourgues and T. Tram, *The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes*, *Journal of Cosmology and Astroparticle Physics* **2011** (2011) 034.
- [5] D. H. Weinberg, J. Miralda-Escude, L. Hernquist and N. Katz, *A Lower Bound on the Cosmic Baryon Density*, *The Astrophysical Journal* **490** (1997) 564.
- [6] D. H. Weinberg, L. Hernquist, N. Katz, R. Croft and J. Miralda-Escudé, *Hubble Flow Broadening of the LY $\alpha$  Forest and its Implications*. eprint: arXiv:astro-ph/9709303, Jan., 1997.
- [7] R. A. C. Croft, D. H. Weinberg, N. Katz and L. Hernquist, *Recovery of the Power Spectrum of Mass Fluctuations from Observations of the Lyman-alpha Forest*, *The Astrophysical Journal* **495** (1998) 44.
- [8] R. Croft, D. Weinberg, N. Katz and L. Hernquist, *Cosmology from the structure of the Lyman-alpha forest*, *arXiv:astro-ph/9801255* (1998) .
- [9] D. H. Weinberg, R. Dav'e, N. Katz and J. A. Kollmeier, *The Lyman-alpha Forest as a Cosmological Tool*, *AIP Conference Proceedings* **666** (2003) 157.
- [10] P. Coles and B. Jones, *A lognormal model for the cosmological mass distribution*, *Monthly Notices of the Royal Astronomical Society* **248** (1991) 1.
- [11] H. Bi and A. F. Davidsen, *Evolution of Structure in the Intergalactic Medium and the Nature of the Ly $\alpha$  Forest*, *The Astrophysical Journal* **479** (1997) 523.
- [12] J.-M. L. Goff, C. Magneville, E. Rollinde, S. Peirani, P. Petitjean, C. Pichon et al., *Simulations of BAO reconstruction with a quasar Lyman-alpha survey*, *Astronomy & Astrophysics* **534** (2011) A135.
- [13] T. Etourneau, J. M. L. Goff et al., *Synthetic quasar spectra with 3D Lyman-alpha correlation*, in preparation (2022).
- [14] T. Etourneau, *Les forêts Lyman alpha du relevé eBOSS : comprendre les fonctions de corrélation et les systématiques*, phdthesis, Université Paris-Saclay, Sept., 2020.
- [15] H. d. M. d. Bourboux, J. Rich, A. Font-Ribera, V. d. S. Agathe, J. Farr, T. Etourneau et al., *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Baryon acoustic oscillations with Lyman- $\alpha$  forests*, *The Astrophysical Journal* **901** (2020) 153.

- [16] N. G. Karaçaylı, A. Font-Ribera and N. Padmanabhan, *Optimal 1D Ly- $\alpha$  Forest Power Spectrum Estimation I: DESI-Lite Spectra*, *Monthly Notices of the Royal Astronomical Society* **497** (2020) 4742.
- [17] N. Palanque-Delabrouille, C. Yèche, A. Borde, J.-M. L. Goff, G. Rossi, M. Viel et al., *The one-dimensional Ly-alpha forest power spectrum from BOSS*, *Astronomy & Astrophysics* **559** (2013) A85.
- [18] A. González-Morales et al., *DESI Lyman-alpha synthetic spectra*, in preparation (2022).
- [19] M. Walther, E. Armengaud, C. Ravoux, N. Palanque-Delabrouille, C. Yèche and Z. Lukić, *Simulating intergalactic gas for DESI-like small scale Lyman $\alpha$  forest observations*, *Journal of Cosmology and Astroparticle Physics* **2021** (2021) 059.
- [20] A. Arinyo-i Prats, J. Miralda-Escudé, M. Viel and R. Cen, *The non-linear power spectrum of the Lyman alpha forest*, *Journal of Cosmology and Astroparticle Physics* **2015** (2015) 017.
- [21] A. S. Almgren, J. B. Bell, M. J. Lijewski, Z. Lukić and E. Van Andel, *Nyx: A MASSIVELY PARALLEL AMR CODE FOR COMPUTATIONAL COSMOLOGY*, *The Astrophysical Journal* **765** (2013) 39.
- [22] Z. Lukić, C. Stark, P. Nugent, M. White, A. Meiksin and A. Almgren, *The Lyman- $\alpha$  forest in optically-thin hydrodynamical simulations*, *arXiv:1406.6361 [astro-ph]* (2015) .
- [23] DESI collaboration, *The DESI Experiment Part I: Science, Targeting, and Survey Design*, *arXiv:1611.00036 [astro-ph]* (2016) .
- [24] DESI collaboration, *The DESI Experiment Part II: Instrument Design*, *arXiv:1611.00037 [astro-ph]* (2016) .
- [25] M. M. Pieri, S. Bonoli, J. Chaves-Montero, I. Paris, M. Fumagalli, J. S. Bolton et al., *WEAVE-QSO: A Massive Intergalactic Medium Survey for the William Herschel Telescope*, *arXiv:1611.09388 [astro-ph]* (2016) .
- [26] P. McDonald, *Toward a Measurement of the Cosmological Geometry at  $z \sim 2$ : Predicting Ly $\alpha$  Forest Correlation in Three Dimensions and the Potential of Future Data Sets*, *The Astrophysical Journal* **585** (2003) 34.
- [27] K. K. Rogers, S. Bird, H. V. Peiris, A. Pontzen, A. Font-Ribera and B. Leistedt, *Correlations in the three-dimensional Lyman-alpha forest contaminated by high column density absorbers*, *Monthly Notices of the Royal Astronomical Society* **476** (2018) 3716.
- [28] S. A. Smee, J. E. Gunn, A. Uomoto, N. Roe, D. Schlegel, C. M. Rockosi et al., *The Multi-object, Fiber-fed Spectrographs for the Sloan Digital Sky Survey and the Baryon Oscillation Spectroscopic Survey*, *The Astronomical Journal* **146** (2013) 32.
- [29] B. Abareshi, J. Aguilar, S. Ahlen, S. Alam, D. M. Alexander, R. Alfarsy et al., *Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument*, Tech. Rep. arXiv:2205.10939, arXiv, May, 2022. 10.48550/arXiv.2205.10939.

- [30] J. E. Gunn, W. A. Siegmund, E. J. Mannery, R. E. Owen, C. L. Hull, R. F. Leger et al., *The 2.5 m Telescope of the Sloan Digital Sky Survey*, *The Astronomical Journal* **131** (2006) 2332.
- [31] D. G. York, *The Sloan Digital Sky Survey: Technical Summary*, *The Astronomical Journal* **120** (2000) 1579.
- [32] K. S. Dawson, J.-P. Kneib, W. J. Percival, S. Alam, F. D. Albareti, S. F. Anderson et al., *The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data*, *The Astronomical Journal* **151** (2016) 44.
- [33] EBOSS collaboration, *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological Implications from two Decades of Spectroscopic Surveys at the Apache Point observatory*, *Physical Review D* **103** (2021) 083533.
- [34] S. Chabanier, N. Palanque-Delabrouille, C. Yèche, J.-M. L. Goff, E. Armengaud, J. Bautista et al., *The one-dimensional power spectrum from the SDSS DR14 Ly $\alpha$  forests*, *Journal of Cosmology and Astroparticle Physics* **2019** (2019) 017.
- [35] C. Pichon, J. Vergely, E. Rollinde, S. Colombi and P. Petitjean, *Inversion of the Lyman  $\alpha$  forest: three-dimensional investigation of the intergalactic medium*, *Monthly Notices of the Royal Astronomical Society* **326** (2001) 597.
- [36] K.-G. Lee, A. Krolewski, M. White, D. Schlegel, P. E. Nugent, J. F. Hennawi et al., *First Data Release of the COSMOS Ly  $\alpha$  Mapping and Tomography Observations: 3D Ly  $\alpha$  Forest Tomography at  $2.05 < z < 2.55$* , *The Astrophysical Journal Supplement Series* **237** (2018) 31.
- [37] B. Horowitz, K.-G. Lee, M. Ata, T. Müller, A. Krolewski, J. X. Prochaska et al., *Second Data Release of the COSMOS Lyman-alpha Mapping and Tomographic Observation: The First 3D Maps of the Large-Scale Cosmic Web at  $2.05 < z < 2.55$* , *arXiv:2109.09660 [astro-ph]* (2021) .
- [38] C. Ravoux, E. Armengaud, M. Walther, T. Etourneau, D. Pomarède, N. Palanque-Delabrouille et al., *A tomographic map of the large-scale matter distribution using the eBOSS Stripe 82 Ly- $\alpha$  forest*, *Journal of Cosmology and Astroparticle Physics* **2020** (2020) 010.
- [39] M. Ozbek, R. A. C. Croft and N. Khandai, *Large-scale 3D mapping of the intergalactic medium using the Lyman  $\alpha$  forest*, *Monthly Notices of the Royal Astronomical Society* **456** (2016) 3610.
- [40] Y.-C. Cai, A. Taylor, J. A. Peacock and N. Padilla, *Redshift-space distortions around voids*, *Monthly Notices of the Royal Astronomical Society* **462** (2016) 2465.
- [41] N. Hamaus, M.-C. Cousinou, A. Pisani, M. Aubert, S. Escoffier and J. Weller, *Multipole analysis of redshift-space distortions around cosmic voids*, *Journal of Cosmology and Astroparticle Physics* **2017** (2017) 014.
- [42] C. Ravoux, E. Armengaud, J. Bautista, J.-M. L. Goff, N. Palanque-Delabrouille, J. Rich et al., *First measurement of the correlation between cosmic voids and the Lyman- $\alpha$  forest*, *arXiv:2203.11045 [astro-ph]* (2022) .

