Neutrinos and dark matter cosmology with the Lyman-α forest : the interplay between large-scale evolution and small-scale baryonic physics

Solène Chabanier

To cite this version:

Solène Chabanier. Neutrinos and dark matter cosmology with the Lyman-α forest : the interplay between large-scale evolution and small-scale baryonic physics. Cosmology and Extra-Galactic Astrophysics [astro-ph.CO]. Université Paris-Saclay, 2020. English. NNT : 2020UPASP034 . tel-03123366

HAL Id: tel-03123366
https://tel.archives-ouvertes.fr/tel-03123366

Submitted on 27 Jan 2021

Neutrinos and dark matter cosmology with the Lyman-α forest: the interplay between large-scale evolution and small-scale baryonic physics

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 576, Particules, Hadrons,
Énergie, Noyau, Instrumentation, Image,
Cosmos et Simulations (PHENIICS)
Spécialités: Astroparticules & Cosmologie

Unités de recherche:
Université Paris-Saclay, CEA, Département d'Astrophysique 91191,
Gif-sur-Yvette, France
Université Paris-Saclay, CEA, Département de Physique des Particules,
91191, Gif-sur-Yvette, France
Référent: Faculté des sciences d'Orsay

Thèse présentée et soutenue à Gif-Sur-Yvette, le 1er octobre 2020, par

Solène CHABANIER

Composition du jury:

Françoise COMBES Professeur, Collège de France	Présidente
Joseph HENNAWI Professeur Associé, Université de Californie, Santa Barbara	Rapporteur & examinateur
Romain TEYSSIER Professeur, Université de Zurich	Rapporteur & examinateur
Jérémy BLAIZOT Astronome Adjoint, Université de Lyon, CRAL	Examinateur
Patrick VALAGEAS Directeur de Recherche, Université Paris-Saclay, CEA IPhT	Examinateur
Frédéric BOURNAUD Directeur de Recherche, Université Paris-Saclay, CEA DAP	Co-Directeur
Nathalie PALANQUE-DELABROUILLE Directrice de Recherche, Université Paris-Saclay, CEA DPhP	Co-Directrice
Yohan DUBOIS Chargé de recherche, Institut d'Astrophysique de Paris	Invité

Thèse de doctorat

NNT: 2020UPASP034

A mes deux grands-mères,
de qui je tiens ma curiosité scientifique et ma volonté d'aboutir ce que j'entreprends.

Remerciements

Tout d'abord, je remercie mes deux directeurs de thèse, Nathalie Palanque-Delabrouille et Frédéric Bournaud. Je vous suis infinement reconnaissante pour cet encadrement exceptionnel dont j'ai bénéficié au cours de ces trois dernières années. Sur le plan scientifique et professionnel, vos expertises complémentaires sur les analyses de données de grands relevés spectroscopiques et sur les méthodes numériques m'ont apporté un éventail de connaissances très poussées sur des sujets transverses et complémentaires qui me passionnent. Je tiens à souligner la pédagogie et la patience dont vous avez fait preuve. Peu importe le sujet, la question ou vos emplois du temps remplis, vous avez toujours répondu présent lorsque j'avais un doute ou que je rencontrais une difficulté. Merci Frédéric pour ta réactivité presque surnaturelle à répondre aux e-mails nuit et jour. Merci pour votre confiance et pour m'avoir octroyée une autonomie dans mon travail, me laissant libre d'explorer les pistes qui m'intéressaient. Merci aussi Nathalie pour m'avoir transmis ton goût pour la diffusion des connaissances, merci pour toutes ces opportunités que tu m'as offertes pour aller à la rencontre du grand public, en particulier pour le festival d'astronomie de Fleurance qui a été une expérience très enrichissante. Ce fut également un cheminement personnel dans lequel vous avez hautement contribué à me faire grandir. Je vous suis très reconnaissante pour votre bienveillance, qui m'a permis de gagner en confiance en moi. Je ne peux pas assez vous remercier pour tout ce que vous m'avez apporté et je suis très fière de l'équipe que nous avons formée durant ces trois années. Malgré ma détermination et ma motivation je n'aurais certainement pas pu accomplir ça sans vous.

Je tiens à remercier très chaleureusement Christophe Yèche, qui a été très présent tout au long de cette thèse et qui a plus qu'activement contribué aux analyses de données et aux analyses cosmologiques. Merci d'avoir toujours gardé ta porte ouverte pour moi quand j'avais besoin d'aide et pour m'avoir toujours prodigué de bons conseils sur la vie en recherche (sans oublier les randonnées chevronnées à faire en Californie). A bientôt j'espère, autour d'une bière à Berkeley ou sur le Half Dome à Yosemite !

Regarding the ensemble of the committe jury I am very proud to have gathered so many experts on the different fields I tackled during my thesis. Thank you to both my referees, Romain Teyssier and Joseph Hennawi, for your attentive reading and fruitful feedbacks on my manuscript. Merci à Françoise Combes pour m'avoir fait l'honneur de présider mon jury.

Je remercie également Anne-Isabelle Etienvre, chef de l'IRFU, ainsi que Gautier, Georges, respectivement chef (lors de la durée de mon contrat) et adjoint-chef du DPhP et Anne chef du DAp, pour avoir permis cette thèse en co-encadrement entre les deux départements. Je réalise la chance que j'ai eu de bénéficier non seulement de deux encadrants de thèse, mais aussi de deux laboratoires multipliant les rencontres, les collaborations ainsi que les champs de connaissances abordés en cosmologie et en astrophysique.

Je remercie les groupes des deux laboratoires du DAp et du DPhP qui m'ont accueillie ces dernières années ; Christophe, Etienne, Eric, Vanina, Jim, Jean-Baptiste et Jean-Marc les permanents du groupe cosmologie du DPhP, merci pour votre bienveillance et pour rendre ce groupe si accueillant. Merci à Emanuele et David, du laboratoire de cosmologie et d'évolution des galaxies du DAp, toutes nos conversations ont toujours été très enrichissantes. David, merci pour ton enthousiasme à partager ta connaissance des sciences, j'ai aussi beaucoup appris de toi.

Je remercie également tous les non-permanent, sans qui le quotidien du laboratoire ne serait pas du tout le même ; merci à Baptiste, mon co-bureau du DAp avec qui j'ai partagé mes deux années de classe préparatoire toujours prêt à m'écouter râler (c'était remarqué et apprécié !). Thank you to Chiara, Boris, Carlos, Menguyan, Benjamin, but also those who already left, Wenjia, Anna and Ivan, thank you for this incredible cultural exchange, for the multiple desserts in the break room that killed so many afternoons of work. Merci Arnaud pour ton sarcasme et pour avoir fait du bureau 41b ce qu'il est aujourd'hui, Emmanuel pour ton enthousiasme et ta positivité à toute épreuve, Corentin pour la bonne humeur et joie de vivre que tu as apporté au 41b, Richard mon co-bureau ces deux dernières années pour toujours avoir fait semblant de m'écouter (même quand ça ne t'intéressait pas), Alex and Michael for your wise advices and support during the post-doc position research. Et bien sur Pauline avec qui j'ai partagé mon bureau durant ma première année et qui m'a toujours bien conseillée même après son départ à Durham, que ce soit sur les projets scientifiques ou de diffusion des sciences, et avec qui j'ai partagé de très bons moments lors du festival d'astronomie de Fleurance.

Merci à toute la collaboration Extreme-Horizon pour avoir mener à bien ce projet, en particulier à Yohan Dubois, véritable pilier dans la réalisation des simulations de la famille Horizon qui a été d'un soutien sans faille lors de la préparation et la mise en oeuvre de mes travaux de simulations numériques. Merci à Damien Chapon et Bruno Thooris, nos collaborateurs du Dedip qui ont produit de magnifiques visualisations de la simulation Extreme-Horizon. Merci également à l'équipe de communication de l'IRFU pour mettre si bien ce projet en avant.

Merci à Sandrine Codis pour ton soutien lors des derniers mois de ma thèse, ton regard sur mon travail et tes encouragements ont eu beaucoup d'impact sur mon moral et ma motivation.

Merci à mes amies d'enfance, Amélie, Joséphine, Claire, Léa et Pauline, qui lorsqu'on s'est rencontrées à l'âge de 10 ans m'entendaient déjà dire que je voulais devenir astrophysicienne. Merci de ne pas m'avoir trouvée bizzare et d'être restées si présentes dans ma vie malgré les différents chemins que nous avons empruntés.

Merci à mes amis, Alex et Sébastien pour ces trois années de colocation durant lesquelles nous avons partagé nos vies, nos bonnes nouvelles et nos coups durs, merci pour votre incroyable soutien, pour m'avoir aidé à décrocher de mon boulot et mon ordinateur, pour ces folles soirées à danser sur du France Gall jusqu'à pas d'heures. Notre quotidien à trois va beaucoup me manquer, sûrement même les dimanches soirs devant le foot.

Je ne serais évidemment pas arrivée là sans ma famille. Merci à toi papa pour m'avoir fait lever la tête vers les étoiles quand j'étais toute petite et à toi maman pour ton regard plein de fierté qui me motive à toujours me surpasser. Merci à mes trois soeurs Audrey, Sabrina et Axelle, et mon beau frère Paul qui ont toujours été là, et seront toujours là, contre vents et marées. Et bien sûr, merci à Saül d'être le rayon de soleil de ma vie.

Finalement, mes dernières pensées sont pour toi Nicolas, avec qui je partage mon quotidien, même de loin, depuis 6 ans maintenant. Les mots ne seront pas suffisants pour te remercier pour tout ce que tu m'apportes, simplement merci d'être qui tu es.

Table des matières

1

Introduction to Modern Cosmology

Contents

COSMOLOGY is a field of Fundamental Science. It is a new science but it aims at answering questions as old as mankind about the our universe ; "Where do we come from ?", "What is the structure and nature of space-time ?" "What is the composition of the universe ?" or "What will happen to us in few billion years ?". Cosmology studies the universe as a whole, its structure, origin, evolution and its different components along with their interactions. Modern cosmology really bloomed in the last century thanks to our capacity to have theories that make predictions and the fact that these predictions can be tested with observations. Recent technological advancements enable to map the sky at an unprecedented level of precision ; as an example, in 1985 the Center of Astrophysics mapped the positions of 1,100 galaxies when the recent state-of-the-art galaxy survey, the Dark Energy Spectroscopic Instrument (DESI), will map the position of 30 million objects in the sky. In particular, the discovery of the General Relativity (GR) theory at the beginning of the last century and the discovery of the expansion of the Universe by Edwind Hubble in 1929 have been true milestones in the constitution of the standard model of cosmology. Nowadays, the current framework of the concordance model is the "Hot Big Bang" model, which states that the universe has expanded from an initial hot and dense state 13.8 billion years ago, that the expansion is still going on and even accelerating today. The dynamics of the universe is well described by the theory of GR applied to an expanding homogeneous and isotropic medium that initially contained tiny density fluctuations that gave rise to the observed large-scale structures growing through gravitational instabilities. This large-scale dynamics is fundamentally connected to the energy density of the different components :

- **Baryonic matter** composes the *visible matter*, or *ordinary matter* from which form stars, interstellar gas and galaxies. It refers to all nuclei and electrons in the universe and represents approximately 5% of the energy budget of the universe.
- **Cold dark matter** is composed by a still mysterious non-baryonic matter necessary to account for the 30% total matter density by, for instance, looking at X-ray emission, at features in the galaxy power spectrum or in the CMB. In particular, with only 5% of the energy budget of the universe composed by baryonic matter, all the structures we observe in the universe would not exist. Thus, most of the matter must not be baryons but must be some new form of matter, dark matter, which is assumed to be non-relativistic, pressureless and represents 25% of the total energy budget. However, its fundamental nature, studied with particle physics experiments and cosmology through its effects on the large-scale structures, is still unkown.
- **Relativistic matter** is composed by relativistic components such as photons or cosmic neutrinos, while they were still relativistic at early times. The temperature of cosmic microwave background photons, i.e. relics from the Big Bang, has been measured very accurately and pins down the relativistic matter contribution to 0.04% today.
- **Dark energy** dominates the energy budget of the universe since it constitutes 69% of the energy content today. It is reponsible for the striking evidences of the observed late-time acceleration of the expansion. Its fundamental nature is also unkown. It can be described with the cosmological constant Λ, from the equations of GR, and the leading candidate for the physical cause of the this Λ would be the vacuum energy. Understanding the late-time cosmic acceleration is a hot topic in cosmology and is the object of many recent ground-based and space-based experiments.

Given this energy content, the concordance model is also called the ΛCDM model. It is elegantly parametrized by only 6 free parameters, called the cosmological parameters, and provides a remarkable fit to much of the existing cosmological data. However, it faces two main challenges :

- It remains a **phenomenological model** because 95% of its energy content is dark and not explained at the fundamental level.
- Theory faces **discrepancies with observations**. Different cosmological probes estimate cosmological parameters that can be in tension at more than the 3σ level and the cold dark matter theory is in tension with astrophysical data.

As such, our universe is far from being understood and many discoveries are still to be made. I feel privileged to have had access to the most advanced data, to large computational power and to invaluable theoretical expertise to contribute to refine our knowledge of the universe.

This Chapter aims at laying down the basics of modern cosmology in order to introduce the fundamental questions my thesis work tackles. I will present in Sec. 1.1 the structure and dynamics of a smooth expanding universe along with the evolution equation of the different cosmic fluid components driving the evolution of the universe structure. In Sec. 1.2, I describe the standard cosmological framework along with its main pillars and observable evidences. Although the standard model has been an incredible success, there are tensions between measurements at the 2-3σ level and also tensions between predictions and observations. Sec. 1.2.4 outlines the two limitations of the ΛCDM model I tackle in this thesis ; the unkown physical parameter related to the masses of neutrinos, whose influence spans the fields of cosmology and particle physics, and the small-scale challenges faced by the cold dark matter model. Finally, in Sec. 1.3 I lay out how to use the heterogeneous matter distribution in the universe to constrain neutrino masses and discriminate between different exotic dark matter models. This chapter is heavily influenced by several cosmology books, in particular FUNDAMENTALS OF COSMOLOGY (Rich, 2001), THE EARLY UNIVERSE (Kolb and Turner, 1990), MODERN COSMOLOGY (Dodelson, 2003), by lectures on cosmology from my master at Imperial College London [1] and by the comprehensive review on neutrino cosmology by Julien Lesgourgues in Lesgourgues et al. (2013).

1.1 A smooth expanding universe

In Sec. 1.1.1 I recall the fundamental concepts of General Relativity that I use as theoretical framework for describing the structure of the universe. In Sec. 1.1.3 I detail the relevant physical laws governing the evolution of the universe and its content in the context of a smooth, isotropic and expanding universe, whose concept is presented in Sec. 1.1.2.

1.1.1 Geometry and General Relativity

In order to study the movement of matter in the universe we need to establish a coordinate system and its metric that assigns to any event one temporal coordinate x^0 and three spatial coordinates, x^1, x^2 and x^3. In any manifold of dimension n, we can define a metric $g_{\mu\nu}$, where $\mu, \nu \in \{0, 1, .., n\}$, to turn coordinate distance into physical distance. Using the Einstein Summation Convention, the distance $\mathrm{d}s$ between $\mathbf{x}$ and $\mathbf{x} + \mathrm{d}\mathbf{x}$ in a manifold is

$$\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu. \tag{1.1}$$

---

[1] `https://www.imperial.ac.uk/study/pg/physics/physics/`

For instance, the metric of a flat space in 3 dimensions can be written in cartesian coordinates

$$g_{\mu\nu} = \begin{pmatrix} +1 & & \\ & +1 & \\ & & +1 \end{pmatrix}, \tag{1.2}$$

so that the line element is

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \tag{1.3}$$

In classical physics, if one observer measures the separation to be δL, then all observers measure the same separation and the same consideration is valid for the passage of time. In Special Relativity and General Relativity (GR) there are no spatial or time conservation but instead there is a combined space-time conservation. By construction, Special Relativity is described by the Minkowski space time with the metric $g_{\mu\nu} = \eta_{\mu\nu}$, which for Cartesian coordinate $(x^0, x^1, x^2, x^3) = (ct, x, y, z)$ is just

$$\eta_{\mu\nu} = \begin{pmatrix} +1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}. \tag{1.4}$$

The great advantage of the metric is that it incorporates gravity. Instead of thinking of gravity as an external force and studying particles into a gravitational field, we can include gravity in the metric and study particles moving freely in a distorted or curved space-time. The Minkowski metric applies only within the context of Special Relativity because it deals with the special case in which space-time is not curved by the presence of mass and energy. Without any gravitational effects, Minkowski space-time is flat and static. When gravity is added, however, the permissible space-times are more interesting. GR theory provides us with the Einstein Field Equations that link the geometry of the universe, encoded in the metric, with the matter and energy contained in the universe, given by

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}. \tag{1.5}$$

Here $G_{\mu\nu}$ is the Einstein tensor, it can be decomposed into the Ricci tensor $R_{\mu\nu}$ which depends on the metric and its derivatives, and R the Ricci scalar, which is the contraction of the metric and the Ricci tensor $g_{\mu\nu} R^{\mu\nu}$. G is Newton's constant and $T_{\mu\nu}$ is the energy-momentum tensor that acts as the source of gravitation, whose components are moments of the distribution function of matter. In GR massive objects distort the space-time curvature through $T_{\mu\nu}$ and other massive objects in the vinicity follow geodesic in a curved space-time. Finally, Λ is the cosmological constant, which was first introduced by Einstein to allow for a static universe, but that he removed afterwards. Nowadays, it is usually replaced by a component (dark energy) with constant energy density. More general formulations of dark energy allow this component to have a redshift-dependent energy density.

1.1.2 The expanding universe

1.1.2.1 The scale factor

The first striking discovery of the expansion of the universe was made in 1929 by Edwin Hubble by showing that extra-galactic nebulae were receeding from each other, as fragments of

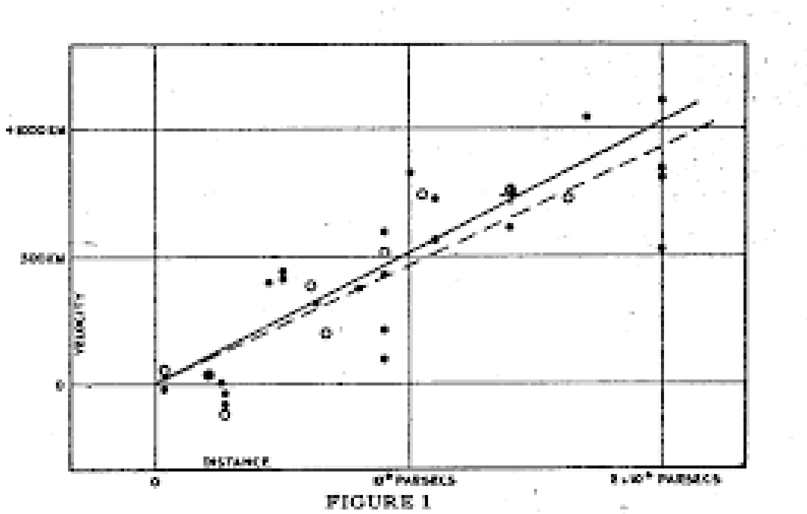

FIGURE 1

FIGURE 1.1 – *The original Hubble diagram from Hubble (1929), i.e. recession velocities as a function of distance. The coefificient of proportionality is known as the Hubble constant H_0.*

a huge explosion Hubble (1929). He combined his measurements of galaxy distances with Vesto Slipher and Milton Humason's measurements of the redshifts (described in the next paragraph, equivalent to the velocity of the galaxy) associated with the galaxies and found that the recession velocity v of galaxies was proportional to their distance d. The original measurement is given in Fig. 1.1. This Hubble law is expressed by the equation $v = H_0 d$, where the proportionality constant, H_0, is named the Hubble constant. Such an expansive trend, currently named the *Hubble flow*, has been confirmed by many other evidences. In particular, the abundance of light elements (hydrogen, helium and lithium) produced during the primordial nucleosynthesis, or the temperature of the relic of the thermal photons emitted during the early universe, can only be explained in the context of an expanding universe. As a direct consequence, it means that early in history distances between us and distant galaxies were smaller than it is today.

It is useful to parametrize the expansion of the universe by a time-dependent function, the scale factor $a(t)$, whose present value is set to one and is increasing with time. We can picture space as a grid, as in Fig. 1.2, which expands uniformly as time evolves. The point coordinates on the grid are constant with time, so their *comoving distance*, $\mathbf{x}$, is constant with time. We can relate comoving distance and true physical distance, $\mathbf{r}$, by

$$\mathbf{r} = a\mathbf{x}. \tag{1.6}$$

The physical velocity can be expanded in two terms :

$$\begin{aligned}\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} &= \frac{\mathrm{d}a}{\mathrm{d}t}\mathbf{x} + a\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} \\ &= \frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t}\mathbf{r} + a\dot{\mathbf{x}} \\ &= H\mathbf{x} + \mathbf{v}_{\mathrm{pec}}.\end{aligned}$$

$\mathbf{v}_{\mathrm{pec}}$ is the galaxy's peculiar velocity, i.e. its velocity with respect to the comoving coordinate system. In the absence of peculiar velocity and in the local universe (i.e. at $t = t_0$), we fall back on the Hubble's law, with the Hubble parameter H, which encodes the logarithmic rate of change in the scale factor at a specific cosmic time, also known as the expansion rate,

$$H(t) = \frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t} \tag{1.7}$$

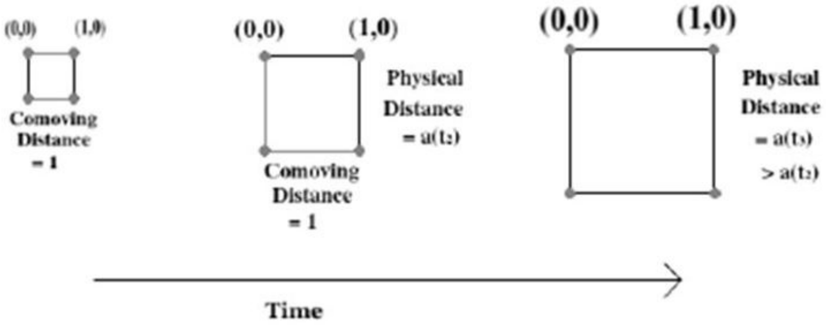

Figure 1.2 – *Evolution of the comoving grid as a function of time. The comoving distance between points on the comoving grid remains constant as the universe expands. The physical distance is proportional to the scale factor, so it gets larger with time. From Dodelson and Efstathiou (2004)*

The Hubble constant at the present time H_0 can be easily extracted from Fig. 1.1, it is simply $H_0 = 500 \text{ km} \cdot \text{sec}^{-1} \cdot \text{Mpc}^{-1}$ almost ten times higher than current estimates since its measurements were dominated by peculiar velocities. In the following for parameters that depend on time, subscript 0 refers to the value of this parameter today. $H(t)$ is often expressed in units of $100 \text{ km} \cdot \text{sec}^{-1} \cdot \text{Mpc}^{-1}$ such that

$$H = 100h \text{ km} \cdot \text{sec}^{-1} \cdot \text{Mpc}^{-1} \tag{1.8}$$

It means that if two galaxies are separated by 1 Mpc today, they receed from each other at the velocity $100h \text{ km} \cdot \text{sec}^{-1}$ The best current estimate of the Hubble constant today is $h \sim 0.7$ but there exists strong tensions at more than 5σ between early-universe inferences, which gives $h \simeq 0.67$ (Planck Collaboration et al., 2015; Addison et al., 2018; Planck Collaboration et al., 2018a) and late-universe measurements, which give $h \simeq 0.74$ (Riess et al., 2016; Abbott et al., 2017; Riess et al., 2019; Freedman et al., 2020).

1.1.2.2 Redshift

One of the main consequence of cosmological expansion is that the wavelength of freely propagating photons increases with the expansion, proportional to the scale factor. The fractional Doppler shift of emitted light resulting from radial motion is encapsulated in the notion of redshift z,

$$1 + z = \frac{\nu_{\text{RF}}}{\nu_{\text{obs}}} = \frac{\lambda_{\text{obs}}}{\lambda_{\text{RF}}}, \tag{1.9}$$

where ν_{RF} and λ_{RF} are the frequency and wavelength emitted in the rest frame, and ν_{obs} and λ_{obs} are the observed ones. In the context of galaxies in the universe, their spectral lines can be redshifted because of three effects :

- the **Doppler shift** is caused by relative peculiar velocity between the source and the observer, it contributes to blue and red Doppler shift
- the **cosmological redshift** is given by

$$1 + z = \frac{a_0}{a}. \tag{1.10}$$

It is a pure consequence of the expansion of the universe and contributes to a red shift

- the **Einstein redshift**, or gravitational redshift, is caused by the difference in magnitude of the gravitational field between the observer and the photon's source if the gravitational field is strong. Except in the close vinicity of strong gravitational fields, such as black holes or neutron stars, this effect is negligible.

From equation (1.10), we see that the redshift can be used as an indicator of time. Today $a = a_0 = 1$ and increases as time increases. So today, $z = 0$ and decreases as time increases. It is equivalently seen as an indicator of distance : the more distant is an object we observe, the more we observe it backward in time due to the finite speed of light, so it will have a high redshift. Whereas a close object, observed in the local universe, is observed almost as it is today, so it will have a low redshift.

1.1.3 Dynamics of the universe

The second axiom we rely on, in addition that GR is correct description of gravity, is the **Comological Principle**. It states that "There is nothing special about our location in the universe" and thus that the universe is homogeneous (constant density) and isotropic (the same in all directions). It is observationally proven that the distribution of matter if *statistically* homogeneous and isotropic to within 10^{-5} on large enough scale, i.e. beyond a few Gpc. Indeed, the world we observe is extremely complex and inhomogeneous with a multitude of fascinating structures but the level of inhomogeneity decreases when we go to greater scales. These heterogeneous structures at small scales arose thanks to tiny fluctuations in the very early universe, which seeded the exceptional initial conditions needed to give a start to the observed world. The characterization of these inhomogeneities will be the subject of Sec. 1.3. In order to understand the framework in which these deviations from smoothness occur, a basic knowledge of the *smooth background universe* is necessary. Also, classical results can be understood in the context of a smooth universe. Thus, I will give in this section the conservation equation for fluid components evolving in a smooth expanding background universe within which GR is the correct theory of gravity. The Einstein Field Equations given by equation (1.5) relate the evolution with time of the cosmic fluid encoded in the stress-energy tensor to the geometry of the manifold. To apply it, we must before get the correct metric and the correct form of the stress-energy tensor for a smooth, isotropic and expanding universe.

1.1.3.1 The Friedmann-Lemaître-Robertson-Walker metric

Applying the Cosmological Principles and the expansion of space to a manifold described by the Minkowski metric, which was shown to be a direct consequence of the postulates of Special Relativity, one can derive the Friedmann-Lemaître-Robertson-Walker (FLRW) metric. In cartesian coordinates, it is expressed as

$$\eta_{\mu\nu} = \begin{pmatrix} +1 & & & \\ & -a^2(t) & & \\ & & -a^2(t) & \\ & & & -a^2(t) \end{pmatrix}. \tag{1.11}$$

In spherical coordinates, the line element ds is

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - a^2(t)\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\,\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2\right), \tag{1.12}$$

where k represents the curvature of the universe : $k = 0$ for a flat universe, $k = 1$ if spherical and $k = -1$ if hyperbolic. By defining the elemental distance $\mathrm{d}\ell^2 = \frac{\mathrm{d}r^2}{1-kr^2} + r^2\,\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2$ in the 3-dimensional hypersurface of spacetime and the *conformal time* such that $\mathrm{d}\tau = c\,\mathrm{d}t/a$, we can further write

$$\mathrm{d}s^2 = a^2(\tau) \times \left(\mathrm{d}\tau^2 - \mathrm{d}\ell^2\right). \tag{1.13}$$

1.1.3.2 The stress-energy tensor

For an ideal fluid with no viscosity whose velocity vector coordinates are denoted u_μ, in a spatially homogeneous and isotropic background, its stress-energy tensor $T_{\mu\nu}$ is expressed as,

$$T_{\mu\nu} = \rho\left[(w+1)u_\nu u_\nu - w g_{\mu\nu}\right], \tag{1.14}$$

where w is the equation of state parameter that relates the pressure p to the density ρ by $p = w\rho c^2$. The perfect fluid approximation is technically not valid in the sense that components of the universe are not fluids. However, in the context of the unperturbated background universe, i.e. in the absence of anisotropic stress, $T_{\mu\nu}$ has the mathematical form of a perfect fluid a posteriori. We consider three different components of the cosmic fluid :

- relativistic matter : $w_\mathrm{r} = 1/3$
- non-relativistic matter with zero pressure : $w_\mathrm{m} = 0$
- dark energy interpreted as a fluid of constant density with negative pressure (such as vacuum or cosmological constant) : $w_\Lambda = -1$.

1.1.3.3 Friedmann's equations

The Einstein's equations (see equation (1.5)) link the Einstein tensor to the source term of curvature, $T_{\mu\nu}$, which results in a set of 10 highly-coupled non-linear second-order differential equations, whose solutions are the component of the metric tensor. These are in general used to derive the metric in regions of space where the wraping of space time is significant, for instance in the vinicity of black holes. In cosmology we work in the opposite way. We postulate the Comological Principle in an expanding space-time and thus assume a FLRW metric , which we use to extract the conservation laws of the cosmic fluid, whose energy density drives the evolution with cosmic time of the scale factor.

For the smooth background, the Einstein tensor $G_{\mu\nu}$ is diagonal, so as is $T_{\mu\nu}$, so the set of 10 coupled equations reduces to only two independent equations. The time component ($\mu = \nu = 0$) of the Einstein's equations (1.5) describes how fast the universe is expanding :

$$\left(\frac{\dot{a}}{a}\right)^2 = H^2 = \frac{8\pi G}{3}\left(\rho + \frac{\Lambda}{8\pi G}\right) - \frac{k}{a^2}. \tag{1.15}$$

The density ρ is in reality the sum of the density of radiation and non-relativistic matter. It is convenient to introduce an energy density for the cosmological constant $\rho_\Lambda = \Lambda/8\pi G$ so that equation(1.15) can be rewritten

$$H^2 = \frac{8\pi G}{3}\rho_\mathrm{tot} - \frac{k}{a^2}, \tag{1.16}$$

where $\rho_\mathrm{tot} = \rho_\mathrm{m} + \rho_\mathrm{r} + \rho_\Lambda$.

Then, by combining the time and the space component of the Einstein's equations, we find an equation for the deceleration of the expansion of the universe :

$$\frac{\ddot{a}}{a} = \frac{-4\pi G}{3c^2}(\rho + 3p). \tag{1.17}$$

Equation (1.16) and equation (1.17) constitute the set of the Friedmann's equations linking the evolution of the scale factor to the energy content of the universe. They are the equivalent of the Poisson and Euler equations for a fluid at rest in a comoving frame.

One can also get the continuity equation by combining the two Friedmann's equations (1.15) and (1.17), which gives :

$$\frac{\dot{\rho}}{\rho} = -3(w+1)\frac{\dot{a}}{a} \Rightarrow \rho(t) = \rho(t_0)a(t)^{-3(w+1)}. \tag{1.18}$$

This conservation law can be immediatly applied to the different constituants to get information about their scaling with expansion :

- For non-relativistic matter : $w = 0 \Rightarrow \rho_{\rm m}(a) \propto a^{-3}$
- For radiatation : $w = 1/3 \Rightarrow \rho_{\rm r}(a) \propto a^{-4}$
- For dark energy : $w = -1 \Rightarrow \rho_\Lambda(a) \propto a^0$

It is convenient to introduce the critical energy density, which is the total energy density of a flat universe today :

$$\rho_{\rm crit} = \frac{3H_0^2}{8\pi G}. \tag{1.19}$$

We also define the density parameter of each component in units of the critical energy density,

$$\Omega_{\rm m,r,\Lambda} = \frac{\rho_{\rm m,r,\Lambda}}{\rho_{\rm crit}}, \tag{1.20}$$

and by analogy we define a density parameter for the curvature term,

$$\Omega_k = -\frac{k^2}{H_0^2}. \tag{1.21}$$

It allows us to rewrite the first Friedmann equation 1.16

$$\left(\frac{H}{H_0}\right)^2 = \Omega_{\rm m,0}a^{-3} + \Omega_{\rm r,0}a^{-4} + \Omega_{\Lambda,0} + \Omega_{k,0}a^{-2}. \tag{1.22}$$

This equation has five parameters. At $t = t_0$, it becomes

$$1 - \Omega_{k,0} = \Omega_{\rm m,0} + \Omega_{\rm r,0} + \Omega_{\Lambda,0}. \tag{1.23}$$

Thus, there are only three out of four Ωs that are independent. We also find that for a flat universe the total energy density is equal to the critical energy density. Due to the change in temperature and pression conditions as the universe expands, the densities of the different components evolved with time and changed their relative ratios such that the universe succesively underwent eras of radiation, matter and dark energy domination. Sec. 1.2.2 presents these different domination periods in the ΛCDM framework.

1.2 The ΛCDM model

The ΛCDM model is a minimal 6-parameter cosmological model starting from a hot Big Bang, for a flat, expanding universe in which gravity is governed by the theory of GR assuming the Cosmological Principle, i.e. the universe is statistically homogeneous and isotropic on large scales. The name refers to the energy content that is dominated by cold dark matter and a cosmologial constant Λ, reponsible for the late-time acceleration of the cosmic expansion. The model also includes a period of rapid acceleration (inflation or a similar process) that occured in the early universe, generating the primordial fluctuations, which seeded large scale structures ; galaxies and clusters which grew through gravitational instabilities during the decelerating matter dominated era. Comparatively recentely, there was a transition to accelerated expansion driven by either a modification of GR or new form of energy. In this section I present its main pillars ; the Hot Big Bang model in Sec. 1.2.1, the universe energy budget in Sec. 1.2.2 and the formation of cosmic structures in Sec. 1.2.3. All data appear to converge toward this model. However it faces challenges and Sec. 1.2.4 outlines the limitations I tackle in this PhD work, related to the masses of neutrinos and small-scale issues faced by the cold dark matter model.

1.2.1 The Hot Big Bang model

The Hot Big Bang (HBB) model postulates that the universe started 13.8 billion years ago when it was much hotter and denser than today. It has been developed in the 1950s by Georges Gamow and Alexander Friedmann. They applied Friedmann's equations and Lemaitre's non static solutions to describe a universe of uniform density and constant spatial curvature. They also extended Lemaitre's idea of a primeval atom by assuming that the early universe was dominated by radiation rather than matter. They introduced the idea of a primordial nucleosynthesis during which the light elements were formed and they also predicted the existence of a relic thermal (blackbody) spectrum of photons, the cosmic microwave background (CMB). Thus, by 1970s, the HBB theory is supported by three major observations :

- the Hubble law which is still the most direct evidence of the expansion of the universe
- the abudance of light elements that were formed during the primordial nucleosynthesis, also called the Big Bang Nucleosynthesis (BBN)
- the black-body radiation of the CMB as a relic of the thermal photons emitted during the early universe

I will here provide the most relevant results for the BBN and the CMB.

Big Bang Nucleosynthesis

When the universe was much hotter and denser, when the temperature of order MeV$/k_B$, there was no neutral atoms or even bound nuclei. The high amount of radiation ensured that any atom or nucleus produced would be immediatly destroyed by a high energy photon. As the universe cooled below the binding energies of light nuclei, light elements began to form. Knowing the conditions of the early universe and the relevant cross-sections, we can calculate the expected primordial abundances of all elements as a function of the baryon to photon density ratio $\eta = \frac{n_b}{n_\gamma}$.

An important quantity to study the thermal history of the universe is the rate of interaction of a reaction Γ, which is the number of reactions per units of Hubble time $t_H = H^{-1}$ and may depends on the temperature T. As long as $\Gamma \gg H$, then each particle will experience interaction and will plausibly remain in thermal equilibrium. However, if $\Gamma \ll H$, the universe expands way faster than the particles interact, and most particles do not interact. They fall-out from equilibrium. This departure from equilibrium is called *freeze-out* as the comoving abundance of a species that has $\Gamma \ll H$ is frozen in the absence of any other interactions. As we go further back in time, the universe gets hotter and hotter as

$$T(z) = T_0(1+z). \tag{1.24}$$

So in the very early universe all particles were relativistic and interactions were strong enough to remain in thermal equilibrium. As the universe cools, particles become non relativistic and possibly fall-out of equilibrium, possibly leaving relics around that we see today. We can divide the thermal history of the universe in four epochs :

- **Thermal equilibrium** ($T \sim 1$ MeV) : The cosmic plasma consists of coupled relativistic (electrons, positrons and photons) and non-relativistic particles (baryons) along with decoupled relativistic particles (neutrinos). Neutrinos decoupled a little above $T =$ 1 MeV and therefore share the same temperature than other relativistic particles and are roughly as abundant but do not couple. The Sakharov conditions states there is a tiny baryons/anti-baryons asymmetry ($(n_b - n_{\bar{b}}) \sim 10^{-10}$), which remains constant throughout the expansion. At $T \sim 1$ MeV all anti-baryons have annihilated away, thus there are many fewer baryons than relativistic particles.
- **Neutron-proton freeze-out** ($T \sim 800$ keV) : the reaction $p + e^- \leftrightarrows n + \nu$ falls out of equilibrium and the neutron-to-proton ratio $n_n/n_p \sim 0.2$ is frozen.
- **Neutron decay** (60 keV $< T <$ 800 keV) : Neutrons are not stable and instead decay into protons $n \to p + e^- + \nu$ with a lifetime of $\sim$ 880s. The duration of this period is about 3 minutes, so about half of the neutrons decay leaving a neutron-to-proton ratio of ~ 0.1.
- **Primordial nucleosynthesis** (30 keV $< T <$ 60 keV) : The remaining neutrons are rapidly consumed in nuclear reactions producing substantial amounts of deuterium, helium ^{3}He and ^{4}He and lithium ^{7}Li. The formation of light nuclei stopped at ^{7}Li because of the absence of stable elements at A = 5,8 and then because the temperature and density conditions were not longer satisfied. The synthesis of heavier elements started up again once stars were formed initiating the stellar nucleosynthesis in their core.

Since we know how matter density evolves with time (it falls as a^{-3}) we can use the measurements of light element abundances in measurements of the baryon density today. Precise measurements of primordial deuterium gives a baryon density Ω_b of at most 5% of the critical density.

Also, I emphasize that the details of nucleosynthesis are heavily influenced by the fact that the reactions involved are eventually too low too keep up with the expansion rate. This feature may also be responsible for the production of dark matter. I will explore this scenario in Sec. 1.3.3.

Cosmic Microwave Background

Elements produced in the BBN remain ionized until the temperature drops well below the the ionization energy of hydrogen. At the epoch of recombination, at $z \sim 1,100$ or about 380,000 years after the Big Bang when $T \sim 0.25$ eV, electrons and protons combine to form neutral hydrogen. Before recombination, photons, electrons and protons are tightly coupled because of Compton and Coulomb scattering. So that the whole formed an ionized plasma called the baryon-photon plasma. After recombination, the mean free path of photons becomes larger than the Hubble horizon so light could cross the universe without scattering. From the time of decoupling until today, these photons were redshifted and formed the *cosmic microwave background* (CMB). Because photons were in equilibrium before recombination and because they are massless, the thermal character of their spectrum is maintened by the expansion. Thus, the CMB we observe today is the redshifted blackbody spectrum of photons produced at the last-scattering surface so that they offer an unvaluable snapshot of the very early universe.

The CMB was discovered in 1964 by Arno Penzias and Robert Wilson who were looking for neutral hydrogen and found a faint and noisy signal while calibrating a microwave antenna of a radio telescope in New Jersey (Penzias and Wilson, 1965a,b). In parallel, the blackbody spectrum radiation had been predicted independently by two teams in the US and in Russia

(Zel'dovich, 1964; Dicke et al., 1965). Since then, the CMB has been measured with very high precision by space-based missions : the Cosmic Background Explorer (COBE) from 1989 to 2007 (Smoot et al., 1992), the Wilkinson Microwave Anisotropies Probe (WMAP) from 2001 to 2007 by the NASA (White, 1999), and the Planck satellite from 2008 to 2013 by the ESA (Planck Collaboration et al., 2018a).

The most important fact we learn from the CMB is that the initial universe was incredibly smooth. However, theoreticians succh as Jim Peebles or Rashid Sunyaev and Yakov Zel'dovich in the 1970s realized there should be structures in the CMB called *anisotropies*, which are the seeds of the large structures we observe today. The temperature fluctuations were first detected by the COBE satellite in 1989 and are now precisely measured to $\delta T/T \sim 10^{-5}$. I will come back to these anisotropies in the context of structure formation in Sec. 1.3. The CMB is today one of the most compelling probe in favor of the Hot Big Bang model and also provides us with extremely precise constraints on cosmological parameters (Planck Collaboration et al., 2015, 2018b).

1.2.2 The cosmic fluid

I will briefly review here the main components of the cosmic fluid and how much energy they contribute to the total energy budget within the ΛCDM model.

1.2.2.1 Radiations

The only relativistic species today in the model are photons from the CMB and massless neutrinos from the cosmic neutrino background. The total relativistic matter population has an energy density today (in units of critical density) of

$$\Omega_{\mathrm{r},0} = 8.24.10^{-5} \tag{1.25}$$

Photons

The temperature of the CMB photons has been measured very precisely by the different CMB experiments at $T = 2.725 \pm 0.002$ K. In the context of a smooth universe, the distribution fonction for photons follows a Bose-Einstein distribution without chemical potential. So the energy density associated with this radiation is

$$\rho_\gamma = 2 \int \frac{p}{e^{p/T} - 1} \frac{\mathrm{d}^3 p}{(2\pi)^3}, \tag{1.26}$$

where p is the photon momentum. Integration and the evolution of temperature with redshift from equation (1.24) give us

$$\rho_\gamma = \frac{\pi^2}{15} T^4 = \frac{2.47 \times 10^{-5}}{h^2 a^4}. \tag{1.27}$$

So that the photon energy density depends on time via the scale factor, with an energy density today of (in units of critical density)

$$\Omega_{\gamma,0} \sim 5.35 \times 10^{-5}. \tag{1.28}$$

Neutrinos

Neutrinos are neutral leptons from the Standard Model of particle physics. They were initially theorized by Wolfgang Pauli in 1930 to solve energy violation of the β radioactive reaction. He dubbed the hypothetical particle **neutrino** as it must be electrically neutral and massless. It was initially thought that neutrinos were undetectable given that they are extremely weakly interacting and are only involved in radioactive decays, such as in the center of stars, nuclear reactor or in the upper atmosphere when particles interact with cosmic rays. But they were finally discovered in 1965 by Clyde Cowan and Frederick Reines when nuclear reactors synthetized antineutrinos by beta decay (Cowan et al., 1956). By analogy with the three generations of electrically charged leptons (the electron, muon and tauon), the Standard Model of particle physics predicts three generations,or *flavours*, of neutrinos : the electronic (ν_e), muonic (ν_μ) and tauic (ν_τ) neutrinos.

Similarly with a relic of CMB photons, the Hot Big Bang model predicts the existence of a sea relic of neutrinos in a number slightly below that of photons. They were once kept in thermal equilibrium with the rest of the cosmic plasma and they had a momentum spectrum with a zero-potential Fermi-Dirac distribution

$$f_\nu(p) = \frac{1}{e^{p/T_\nu}+1}, \tag{1.29}$$

which gives us the following equation for the energy density of each neutrino :

$$\rho_\nu = 2\int \frac{\sqrt{p^2+m_\nu^2}}{e^{p/T_\nu}+1}\frac{\mathrm{d}^3p}{(2\pi)^3} \tag{1.30}$$

To get their energy density today, we need to relate their temperature to that of photons, which we do observe on the opposite of neutrinos. Also, because neutrino masses are very small compared to their temperature, we can use the approximation $m_\nu = 0$. At later time, when $T \sim 1$ MeV the rate of weak interactions ($e^- + e^+ \leftrightarrows \nu_e + \bar{\nu_e}$) falls below the Hubble time and each generation of neutrino freezes-out of thermal equilibrium. However, their distribution remains Fermi-Dirac, with their temperature simply falling as $1/a$ as given by the temperature evolution equation (1.24). Shortly after, at $T \sim 0.5$ eV, the annihilation of electrons and positrons acts as a source of additional heating of photons. Since neutrinos lost contact slightly before this annihilation they did not inherit of the associated energy and are cooler than photons by a factor

$$\frac{T_\nu}{T_\gamma} = \left(\frac{4}{11}\right)^3. \tag{1.31}$$

Today the cosmic neutrino background has cooled down to $T_{\nu,0} = \left(\frac{4}{11}\right)^3 T_{\gamma,0} = 1.9525$ K. By injecting this temperature in equation (1.30), we get the total massless neutrino energy density today is (in units of critical density)

$$\Omega_{\nu,0} = 0.97\times10^{-5}\times N_\nu, \tag{1.32}$$

where $N_\nu = 3$ is the number of neutrinos. However, the neutrino decoupling is not an instantaneous process, such that a small portion of neutrinos are also affected by the electron-positon annihilation slightly modifying the neutrino temperature. We correct for this effect by introducing the *effective number of neutrino species*, $N_{\text{eff}} = 3.046$ which would be 3 in the aproximation of instantaneous decoupling of neutrinos from photons. The CMB temperature anisotropies (Planck

Collaboration et al., 2018a) constrain the effective number of stable, relativistic species in thermal equilibrium in the early universe to

$$N_{\text{eff}} = 2.99 \pm 0.17. \tag{1.33}$$

So the correct expression of the neutrino energy density is

$$\frac{\rho_\nu}{\rho_\gamma} = \frac{7}{8}\left(\frac{4}{11}\right)^3 N_{\text{eff}}. \tag{1.34}$$

The cosmic neutrino background is so cold and so weakly interacting that direct detection experiments are very challenging. Measurements of the light element abundances are sensitive to the neutrino density and have already put strong constraints but are somewhat very model dependent. An additional probe of the existence of this neutrino background comes from their characteristic effet on perturbations in the primordial plasma. Since neutrinos travel at the speed of light, they travel significantly faster than sound waves in the primordial baryon-photon fluid, which travels at $\sim c/\sqrt{3}$. Thus, relativistic neutrinos induce perturbations beyond the sound horizon of the baryon acoustic oscillations and the gravitational influence of this supersonic propagation induces a shift in the phase of the accoustic oscillations that cannot be mimicked by other properties of the plasma. This phase shift has been recently detected in the CMB (Follin et al., 2015) and in the distribution of BOSS galaxies (Baumann et al., 2019). They both add to the robustess of cosmological evidences of this cosmic neutrino background.

1.2.2.2 Non-relativistic matter

Non-relativistic matter is decomposed in two components ; the baryons and dark matter. Its total energy density (in units of critical density) is

$$\Omega_{\text{m},0} \simeq 0.3. \tag{1.35}$$

Baryons

Baryons represent all nuclei and electrons. This is technically not correct since electrons are leptons but given that nuclei contain the dominant fraction of the mass by several orders of magnitude all the mass is in baryons. Baryons constitute the visible matter : gas, stars, dust and planets. There are many ways to estimate the baryon density and they all agree quite well. We can look at baryons into galaxies today, but this method tends to under estimate $\Omega_{b,0}$ because most of baryons are too cold or too diffuse to be readily visible. We can also estimate it by studying the neutral hydrogen absorption features in the spectra of distant background objects, which provides us with information on the quantity of hydrogen in the intergalactic medium that do not emit light (Péroux et al., 2003; Noterdaeme et al., 2012). The light element abundances are also very dependent on the baryon density at that time, which we can extrapolate to today since we know that the comoving density scale as a^{-3} (Burles et al., 2001). Finally the CMB also provides very strong constraints on the baryon density via the *baryon accoustic oscillation* (BAO) mechanism that I will describe in Sec. 1.3 (Planck Collaboration et al., 2015, 2018c). All these methods pin down the baryon density to

$$\Omega_{b,0}h^2 = 0.022. \tag{1.36}$$

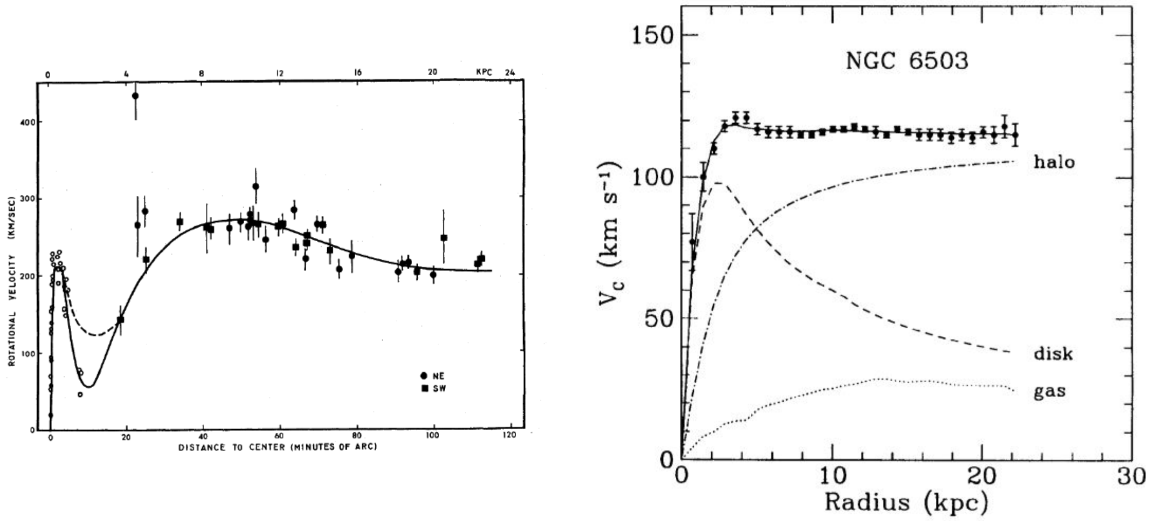

FIGURE 1.3 – *Rotation curves of spiral galaxies* ***Left :*** *BThe original rotation curve of Andromega by Rubin and Ford (Rubin and Ford, 1970).* ***Right :*** *The neat curve of NGC6503 with predictions of what is expected for the visible disk. From (Begeman et al., 1991).*

Cold dark matter

The first compelling evidence for dark matter was made by the astronomer Fritz Zwicky, in the 1930s. By observing the Coma cluster of galaxies he noted that the dispersion in the radial velocity of the cluster's galaxies was very large, around 1,000 km $\cdot$ s^{-1} (Oort, 1932; Zwicky, 1933). The inferred mass of visible matter formed by gas and stars could not provide enough gravitational attraction to hold the cluster together and he dubbed this "missing mass" necessary to hold the galaxy together dark matter. Other compelling evidences of departure from the predictions of Newtonian gravity at the galactic scale were made by Vera Rubin and W. Kent Ford in 1970. They were the first to perform a precise measurement of the rotation curve of the Andromeda galaxy (M31) tracing about 70 hydrogen clouds. They determined the curve to be rather flat far from the center (out to $\sim$ 22 kpc), as illustrated in Fig. 1.3. The constant velocity at high radii can only be explained in the presence of a non-visible massive halo around the galaxy, called a dark matter halo.

Since then, dark matter evidences have emerged at all scales. The presence of dark matter on the scale of galaxy clusters comes from the observations of a pair of colliding clusters known as the 'bullet cluster' (Clowe et al., 2006). Fig. 1.4 shows on the left panel the image of these colliding clusters ; the gas heats and radiates X-rays shown in pink while the blue cloud shows the region where most of the mass lies, as deduced by gravitational lensing (an example of strong gravitational lensing is shown on the right panel). The special feature of this system is that visible and dark matter are spatially separated. Before colliding, the two systems were separated within each dark and visible matter mixed together. When the two systems collided 150 million years ago, visible matter that interacts significantly with itself experienced a collisional shock wave and got heated. However, dark matter only weakly interacts with itself and ordinary matter, so dark matter clouds simply passed through each other without collisions. Since then, many similar systems have been discovered, the authors in Harvey et al. (2015) reports 72 of them.

We also find striking evidences of dark matter on the largest observable scales, from the large-scale structures and the CMB. Indeed, without dark matter the universe would not appear as it is today, the galaxies could not be distributed this way and the CMB temperature anisotropies

FIGURE 1.4 – ***Left*** *: Composite optical image of the Bullet Cluster (credit : NASA/ STScI ; Magellan / U. of Arizona).* ***Right*** *: Gravitational lensing manifest near the 0024+1654 cluster (credit : HST), distorting the light rays from a background galaxy, shown as the stretched blue streaks.*

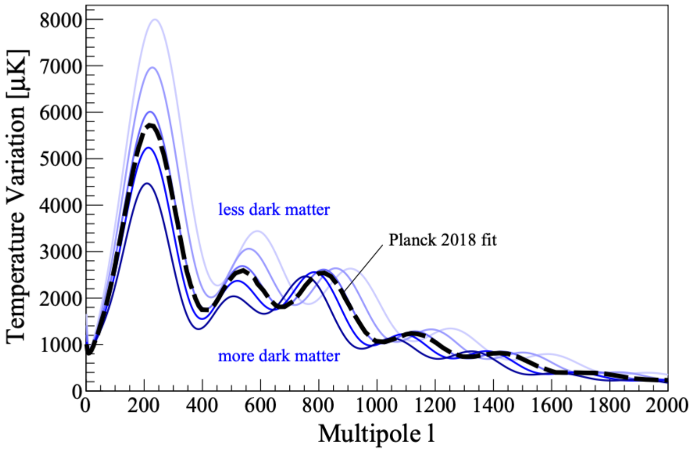

FIGURE 1.5 – *Variation of the CMB power spectrum when increasing or decreasing dark matter energy density and keeping the other parameters constant. The dashed line shows the best-fit Planck 2018 release (Planck Collaboration et al., 2018b). From Schumann (2019).*

would also look different, as illustrated in Fig. 1.5. Thanks to dark matter, the universe went from an almost perfect smoothness state with small perturbations, to a state with multiple structures at all scales. I will present in Sec. 1.3 the evolution of the different component perturbations and we will see that ordinary matter perturbations alone does not allow the construction of cosmic structures because all interactions it experience impede its clustering, at least at early times.

We therefore have a significant number of evidences telling us that the total matter density is about five times larger than that of the visible matter : so most of the matter in the universe must not be baryons but some new form of matter, the dark matter that interacts very weakly with ordinary matter except via gravitation. The dark matter density today (in units of critical density) is

$$\Omega_{\mathrm{DM},0} = 0.258 \pm 0.0020, \tag{1.37}$$

meaning dark matter constitutes about 84% of the total matter content and 26% of the total matter-energy content.

The standard model of cosmology is based on *cold, collisionless* and *stable* dark matter with adiabatic inhomogeneities. Cold means that dark matter behaves as a non-relativistic fluid, i.e. the typical momentum of dark matter particles p, is much smaller than their mass, $p \ll m$, or equivalently their typical velocity v is much smaller than the speed of light $v \ll c$. Thus, if dark matter is made of thermalized particles, this implies that they must be heavier than a few keV. Collisionless means that the interaction among dark matter particles or between dark matter and other particles are small enough to not be detected. This is what significantly differentiates dark matter from ordinary matter, which is sensitive to many interactions, notably electromagnetic. However, this does not mean dark matter does not interact at all, it only means the interactions are very weak w.r.t. our sensitivity threshold today. Stable means dark matter is present since the early moments of the universe and has not disappeared yet. If dark matter does decay, its half-life is much longer than the age of the universe so that the cosmological effects are negligible. Finally, the adiabatic inhomogeneities of dark matter means it has the same primordial density fluctuations as other particles.

There are a large number of candidate particles for cold dark matter. The most popular one since 1980s were WIMPs (Weakly interacting Massive Particles). They arise naturally in various theories beyond the Standard Model of Particle Physics, e.g. in supersymmetric theories (Nilles, 1984; Lahanas et al., 2007; Peskin, 2015) or in theories with extra spacetime dimensions (Hooper and Profumo, 2007). They constitute a rather model-independent generic class of dark matter with masses ranging between 1 to 10^2 Gev·c^2 and interaction cross-sections from 10^{-41} to 10^{-51} cm^{-2}, so they would behave as *cold relics*. The interesting fact that without a fine-tuning of parameters, a WIMP could produce the correct amount of dark matter is called the "WIMP miracle". Axions are also excellent candidates. These are scalar field particles predicted to solve a problem in quantum chromodynamics (Ringwald, 2012). However, despite all effort, so far, no convincing detections have been made but the search goes on. It is worth noting that at the time of writing this thesis the XENON1T experiment, designed to look for WIMPS, detected suspiscious excess of events over low background below 7 keV, rising toward lower values (Aprile et al., 2020). This signal could come from contamination inside the experiment, but also fit the profile of hypothetical dark matter particles such as axions or non-standard neutrinos with large magnetic moment. Future experiments are already on track to confirm this finding and discriminate among the different plausible explanations. I will consider in this thesis the plausibility of a model of warm dark matter constituted by sterile neutrinos or thermal relics in the context of structure formation. I will precisely describe the models in Sec. 1.3.3.

1.2.2.3 Dark Energy

For a homogeneous universe filled with matter or radiation, GR predicts that the cosmic expansion will slow down over time. However, in the late 1990s two independent studies of distant supernovae found that the expansion has accelerated during the last 5 billion years (Riess et al., 1998; Perlmutter, 1999). Also, the strong evidences for spatial flatness and improving measurements of the Hubble constant provided independent argument for an energetically dominant accelerating component (de Bernardis et al., 2000; Hanany et al., 2000; Mould et al., 2000). The fact that on cosmic scales gravity repels has been measured by many independent probes, e.g. larger and better calibrated supernova sample, high-precision CMB data, the baryon acoustic scale in galaxy clustering, weak lensing measurements or the abundance of massive clusters in

X-ray surveys. Explaning all this data simultaneously requires an accelerating universe which is accounted for by a still-unknown dark energy.

A cosmological constant is the simplest solution to this cosmic acceleration. In his Einstein Field Equations, Einstein initially introduced a cosmological constant to make the universe static, but it was removed by the scientific community when Hubble discovered direct evidence for the expansion of the universe (Hubble, 1929). While Einstein introduced it as a modification of the curvature side, its is now more common to interpret Λ as en energy component constant in space and time, thus on the stress-energy side. It can be reasonnably intepreted as the gravitational signature of the quantuum vacuum energy. Several cosmological probes either measure independantly or infer

$$\Omega_\Lambda = 0.69. \tag{1.38}$$

But there is a huge problem of magnitude since estimations of the vacuum energy are about 10^{120} orders of magnitude larger than the observed value for the cosmological constant (Weinberg, 1989). Some extensions of this theory consider dynamical dark energy, whose energy varies with time (Ratra and Peebles, 1988; Ferreira and Joyce, 1997). The alternative to introducing a new energy component is to modify GR on cosmological scales which can alter the relation between the expansion history and the growth of matter clustering (Capozziello and Lambiase, 2002; Carroll, 2004; Dvali, 2004).

1.2.2.4 Domination periods

In the Hot Big-Bang model scenario, the universe formed 13.8 billion years ago and its expansion rate is driven by the energy density of the cosmological fluid according through the Friedmann's equations (1.16) and (1.17). The cosmological fluid consists of three components : radiation, matter and a cosmological constant, each having the equation of state I recall in Tab. 1.1. Because the energy densities of the three components evolve distinctly with the scale factor, we can identify three epochs of the history of the universe during which the radiation, matter or cosmological constant components dominates the others.

With today's values given in equations (1.25), (1.35) and (1.38), we can compute that the universe starts with a period of radiation domination up to $z_{eq} \simeq 3,400$, following by a period of non-relativistic matter domination up to $z_\Lambda \simeq 0.3$. Fig. 1.6 shows the evolution of Ω_r, Ω_m and Ω_Λ with the scale factor. Tab. 1.1 also gives the evolution of the scale factor with proper time in the three periods of domination.

component	equation of state	energy density	scale factor	z-domination period
radiation	$w_\mathrm{r} = 1/3$	$\rho_\mathrm{r} \propto a^{-4}$	$a(\tau) \propto \tau$	$-\infty$ to 3,400
matter	$w_\mathrm{m} = 0$	$\rho_\mathrm{m} \propto a^{-3}$	$a(\tau) \propto \tau^2$	3,400 to 0.3
dark energy	$w_\Lambda =$ -1	$\rho_\Lambda \propto a^0$	$a(\tau) \propto -\tau-1$	0.3 to today

TABLE 1.1 – *Equation of state and evolution of their density as a function of scale factor along with the evolution of the scale factor as a function of conformal time during their domination epoch.*

1.2.3 Inflation and structure formation

The cosmological principle states that the universe is homogeneous and isotropic. This implies that a region at least as large as our present Hubble volume is smooth, which is statistically

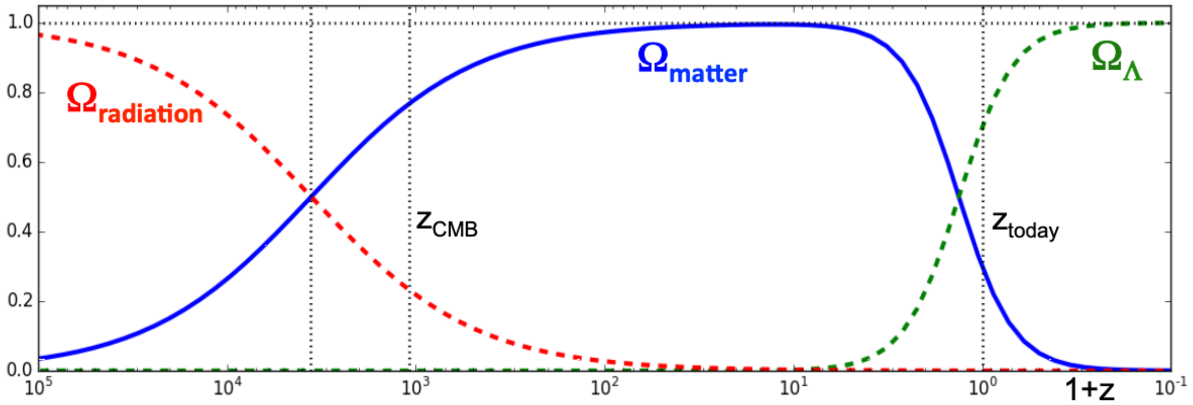

FIGURE 1.6 – *Energy density of radiation (red), non-relativistic matter (blue) and dark energy (green) as a function redshift. vertical dashed lindes indicate the redshift of radiaiton-matter equality, the recombination (z_{CMB}) and today.*

verified on very large scales. However, the level of inhomogeneity increases as scale decreases and we observe heterogeneous structures with clusters, voids, filaments, galaxies, wall of galaxies, stars and planets. Also, if we look at the million of galaxies mapped in the Sloan Digital Sky Survey for instance, we see that galaxy positions are not randomly distributed : the universe has structure on large scales. To understand this structure, we must allow for deviations from smoothness.

Initial tiny perturbations, detected as temperature fluctuations in the CMB, are the seeds of the galaxies we see today. The origin of these initial conditions is not explained by the model and is not understood yet. The best candidate is the amplification and freezing of quantum fluctuations during inflation, the short phase of exponential growth that the universe underwent shortly after the Big Bang which would explain the homogeneity and flatness in the universe. However, we are still not sure it is the correct mechanism. The detection of gravitational wave from cosmic inflation in the CMB would be compelling evidence for inflation.

As time evolves over-dense regions contract and are amplified by gravitational instabilities. They eventually collapse to form the structures we observe today. At the same time, under-dense regions get more and more empty, and constitute cosmic voids. Gravitational instabilities result from the competition between gravity, which tends to amplify local density fluctuations, and pressure (from baryons) and expansion which suppress them. There are two hierarchical scenarios for structure and galaxy formation ; the *bottom-up scenario* in which low-mass structures form first before merging into larger structures, and the *top-down scenario*, in which largest structures form first thereafter subdividing into clusters, groups, and galaxies. The bottom-up scenario is an important prediction of the cold dark matter model. If on the opposite we consider a hot dark matter model, i.e. with weakly interacting particles with large velocity dispersion, small-scale structures and galaxies end up forming significantly later than in the cold dark matter case, which is in strong disagreement with observations.

Because dark matter constitutes the essential of the total matter and is sensitive only to gravity, it drives the collapse of structures. After collapse of dark-matter halos, primordial gas collapse following the gravitational potential of dark matter and settles at the virial equilibrium in the dark matter potential well. The difference in the subsequent evolution of gas and dark matter is that gas can cool radiatively. Thus, it then falls toward the center of the dark matter halo and settles in rotationally supported disks.

The distribution of matter in the universe thus results from the gravitational amplifications of tiny inhomogeneities in the early universe. I will present the equations governing perturbations around a smooth background in Sec. 1.3 along with the statistical tools to learn about cosmology by studying these large-scale structures.

1.2.4 Limitations

The theory encompassing a statistically homogeneous and isotropic universe, the Hot Big Bang model and the evolution of structures seeded by small initial fluctuations that grow through gravitational instabilities driven by cold dark matter, constitutes the ΛCDM model toward wich all data appear to converge. It is an impressively successful framework for predicting and explaining many independent observations, from the CMB to large-scale structures.

However, cosmology is far from having reached its final goal since the concordance model still faces many challenges. First, it remains a phenomenological model since the actual physics of dark matter and dark energy are still not understood. In addition, we have now entered the era of *precision cosmology* meaning that the precision of measurement of cosmological parameters have reached the percent level. Thus, even if all data converge toward the ΛCDM model, some independent probes now exhibit tension at the 5σ level. A good example of a recent tension is the H_0 tension between early-universe and late-universe measurements, presented in Sec. 1.1.2. A substantial discordance is also found on structure growth rate measurements between weak lensing measurements (Troxel et al., 2017; Abbott et al., 2018; Shan et al., 2018) and ΛCDM extrapolation of CMB measurements (Planck Collaboration et al., 2015, 2018b).

These divergences could be the results of unidentified biases in data analysis or could reflect the need of new physics beyond the standard model. Even if these deviations appear to be small, they are significant enough to have the power to discriminate different cosmological models. In this context, my PhD work tackles some of these fundamental issues : massive neutrinos and the small-scale challenges of the cold dark matter scenario, which I describe in the following paragraphs.

1.2.4.1 Massive neutrinos

Cosmic neutrinos are part of the cosmic fluid and the Standard model of Particle Physics predicts *massless* neutrinos. However, the discovery of the neutrino flavour oscillation in 2000 (Fukuda et al., 1998; Kajita, 1999; Ahmad et al., 2001), meaning the observable transformation of a neutrino from one flavour to another flavour, has been a major breakthrough for particle physics since it is only possible if neutrinos are massive. Predicted in 1957 by Pontecorvo (1968), these oscillations have first been hinted by the solar neutrino problem from the Homestak experiment that reported a deficit in the flux of solar neutrinos with respect to standard model predictions (Davis et al., 1968). If neutrinos were massless, they would propagate at the speed of light and experience infinite time dilatation. Since neutrino lepton flavours oscillate, their time dilatation cannot be infinite and they therefore cannot travel at the speed of light, which means at least one neutrino must have non-zero mass. Thus, it indicates a serious failure of the standard model.

To understand this phenomenon we need to introduce the concept of *flavour eigenstates* and *mass eigenstates*. The neutrino flavour identifies the neutral particle associated by weak interactions to the charged leptons with given flavor, i.e. the electron with electronic neutrino ν_e, muon with muonic neutrino ν_ν and tauon with tauique neutrino ν_τ. For instance, if we consider the β^+ decay of ^{37}Ar we associate the electronic neutrino, to the electron with $\mathrm{e}^- +^{37}\mathrm{Ar} \to \nu_\mathrm{e} +^{37}\mathrm{Cl}$. The association works such that the flavour is conserved for a given interaction. These flavour eigenstates are each a different superposition of the three neutrino mass eigenstates ν_1, ν_2 and ν_3. Neutrinos are emitted and absorbed in weak interactions in their flavour eigenstate while they propagate as mass eigenstate. The neutrino flavour eigenstate can be written as a linear superposition of the mass eigenstates,

$$\nu_\ell = \sum_{i=1}^{N} U_{\ell i}\nu_i \quad \text{with} \quad \begin{cases} \ell = \mathrm{e}, \mu, \tau & [\text{ flavor }] \\ i = 1, 2, 3 & [\text{mass}] \end{cases} \tag{1.39}$$

where $U_{\ell i}$ is the mass-flavour mixing matrix, or the Pontercorvo-Maki-Nakagawa-Sakata (PMNS) matrix (Pontecorvo, 1968). It depends on three active mixing angles, a CP violating phase, which has lately been detected at 3σ (Abe et al., 2019), and the three squared mass differences Δm^2_{21}, Δm^2_{32} and Δm^2_{31}. Another unknown of neutrinos properties is their Dirac or Marojana nature. As the neutrino superposition propagates through space, the phase of the three mass eigenstates advance at different rates if the neutrino masses are not identical. This results in a modification of the superposition mixture of the mass eigenstate thus it corresponds to a different mixture of flavor states. For instance, a neutrino produced as an electronic neutrino will be observed as a mixture of electronic, muonic and tauic neutrinos later in time. Therefore, we must have at least one massive neutrino to allow for the observed neutrino oscillation. There are many well-motivated extended models where neutrinos acquire a mass. It is fundamental to measure precisely neutrino masses to be able to discriminate between these models.

From a point of view of cosmologist, we have now reached a high enough level of precision in our measurements to be significantly sensitive to the impact of massive neutrinos, which alter the history of the universe and structure formation compared to a massless neutrino ΛCDM model. Indeed, neutrino particle physics experiments show that their individual mass is quite small but since they are extremely abundant (~ 113 relic neutrinos per cm^3 for each species) the consequences for the universe evolution can be profound. In particular, it modifies the redshift of matter-radiation equality (which is the beginning of structure formation) since at least two neutrino states have a large enough mass for being non-relativistic today, thus making up a small fraction of the dark matter of the universe and suppressing small-scale cosmic structures. As an extreme case, with a low Hubble parameter $h \sim 0.5$, an average mass of neutrino of ~ 8 eV would be sufficient to close the universe ($k < 0$). Measuring their mass is therefore a necessity to improve our picture of the universe.

Unfortunately, neutrino oscillation experiments are insensitive to their absolute mass scale and only constrain the squared mass differences : Δm^2_{21}, Δm_{32} and Δm^2_{31}. The authors in Ahmad et al. (2001) established in the Sudbury Neutrino Observatory (SNO) experiment, which studies solar neutrino oscillations, that $\Delta m^2_{21} = (7.53 \pm 0.18) \times 10^{-5}\mathrm{eV}^2 > 0$. Atmospheric neutrino oscillation studies measured $|\Delta m^2_{32}| \simeq |\Delta m^2_{31}| = (2.44 \pm 0.06) \times 10^{-3}\mathrm{eV}^2$. However it is still unclear today wether $m^2_3 > m^2_2$ or $m^2_3 < m^2_1$, each case being known as the **normal hierarchy** (NH) and **inverted hierarchy** (IH) respectively, illustrated in Fig. 1.7. In the NH case, there are two light states, split by δm^2, and a heavier one separated from the two other by Δm^2, the minimum total mass is thus $\sum m_\nu = 0.06$ eV. In the IH case, there are two heavy states and one lighter, the minimum total mass is $\sum m_\nu = 0.10$ eV.

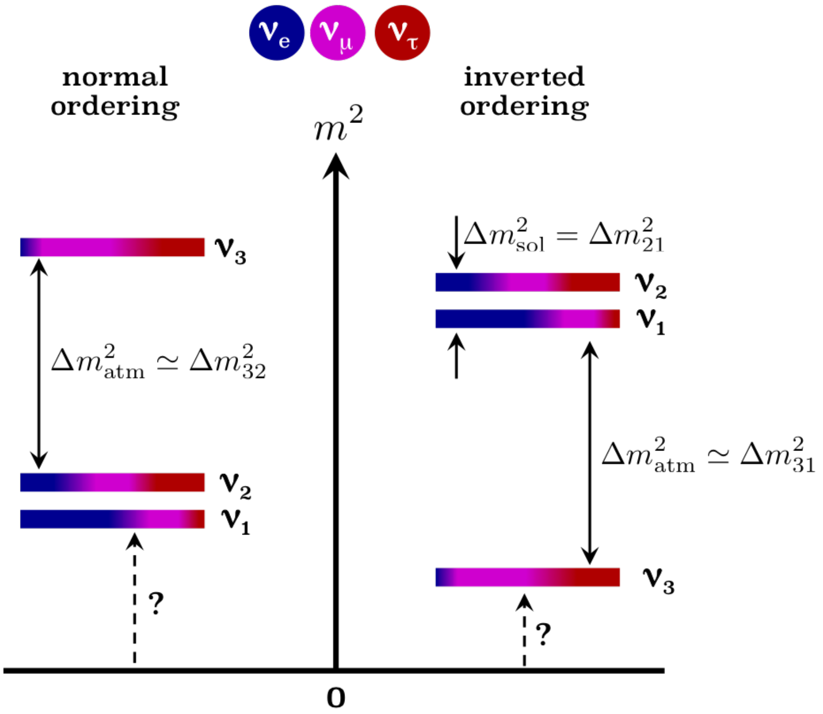

FIGURE 1.7 – *Normal (left) and inverted (right) neutrino hierarchy. Credits : Julien Baur*

To determine the asbolute scale of neutrino masses one can go about two ways : measure either

- the lightest mass state which can be measured with the β decay from tritium $^3_1$H, attempted by experiments such as KATRIN (Weinheimer and KATR. I. N. Collaboration, 2002).
- the sum of all three neutrinos $\sum m_\nu = m_1 + m_2 + m_3$ which is attempted by cosmological probes such as the CMB, or as is the case in this PhD work, the Lyman-α forest. The sum of the neutrino masses is restricted to the approximate range

$$0.06(0.1)\text{eV} \leq \sum m_\nu \leq 3\text{eV}, \qquad (1.40)$$

where the upper limit comes exclusively from tritium beta decay results (recently improved from 6eV (Aker et al., 2019)) and the lower limit reflects the minimum value for the normal (inverted) hierarchy.

I will discuss with precision in Sec. 1.3.2 how massive neutrinos affect the formation of structures and how we can efficiently improve our knowledge of neutrinos with cosmology.

1.2.4.2 Small-scale challenges for cold dark matter

We now know that dark matter is the backbone of structure formation since it represents the essential of the matter energy budget. The cold dark matter model, thus weakly interacting particles with very small velocity dispersion, is incredibly successful at explaining the formation of cosmic structures over an enormous span of redshifts. However, the blooming of dark-matter only cosmological simulations (N-body simulations) since the 90' enabled to reveal tensions with observations on scales smaller than 1 Mpc $\cdot$ h^{-1}. There are three classical problems associated with the small-scale predictions for dark matter in the ΛCDM framework.

- **The Missing Satellite problem :** The highest-resolution N-body simulations have demonstrated that dark matter clumps exist at all resolved masses. Thus, for a Milky-Way like galaxy, we expect thousands of subhalos with mass in principle large enough ($M_{\rm halo} \sim 10^6$ M$_\odot$) to allow molecular cooling which would make them visible. However, we only have observed ~ 30 small-mass satellites around the Milky-Way. Even if there is real hope that future surveys will reveal ultra-faint dwarf galaxies it seems unlikely there are thousands of undiscovered dwarf galaxies.
- **The Cusp-Core problem :** The central regions of dark-matter dominated galaxies as predicted from numerical simulations have "cuspy" distributions, with density increasing steeply at small radii, whereas the rotation curves of observed galaxies suggest that they have flat central dark matter density profiles.
- **The Too-Big-to-Fail problem :** Similarly with the Missing Satellite problem, the local universe contains too few galaxies within massive dark-matter halos ($M_{\rm halo} \sim 10^{10}$ M$_\odot$). Halos of this mass are generally believed to be too massive to have failed to form stars. Thus, the fact they are missing is hard to understand given that they should be visible.

One popular solution is to invoke baryonic effects. Indeed, the initial predictions from which arose the small-scale challenges were performed by means of N-body simulations, i.e. only evolving dark matter. Baryonic feedbacks have often been believed to be the key to solve these issues. Indeed, for the Missing Satellite problem, strong galactic winds would be powerful enough on such a range of mass to efficiently supress gas accretion and star formation, so that these small-mass halos would remain dark. For the cusp-core problem, hydrodynamical simulations (simulations evolving dark matter and baryons) have shown that it is posisble for baryonic feedbacks to produce core-like density profile. If galaxies form enough stars there will be enough energy from supernovae to redistribute dark matter and create cores. If too many baryons end up in stars, however, the excess central mass can compensate and drag dark matter back in. While some hydrodynamical simulations show that baryonic effects can help *alleviate* the tensions they do not appear to fully solve them, with the resolution and physical models implemented today. It is important to stress that the small-scale issues presented above rely on theoretical predictions of simulations. Maybe these issues would not exist with an unlimited computational power or if physical models of galaxy formation were fully understood. As I will present in Chapter 5 numerical simulations are the most accurate tools we have to make theoretical predictions but they can also sometimes present large variations in galaxy properties because of our lack of knowledge in their formation processes.

That being said, these small-scale challenges have been persistent along the years, in spite of the increasing numerical power. In addition, we still have no clue of the fundamental nature of dark matter particle since particle physics experiments have not discovered yet new particles beyond the standard model of particle physics with appropriate characteristics. Instead, all these problems along with the absence of detection of dark matter particles could indicate a failure of the cold dark matter hypothesis. Many theoretical models have emerged to concile large and small scales predictions with observations, with for instance wark dark matter model constituted by sterile neutrinos (Merle, 2013), self-interacting dark matter with large cross-section (Spergel and Steinhardt, 2000; Rocha et al., 2013), decaying dark matter with a small mass splitting (Wang et al., 2013) or extremely light bosons (Hu et al., 2000; Marsh, 2016). In my thesis, I have tested the plausibility of a warm dark matter model. I will describe in Sec. 1.3.3 the impact of such a scenario on structure formation.

1.3 Structures in the universe

In the previous section, I presented the basics of the standard ΛCDM cosmological model along with two of its limitations : the masses of neutrinos and the small-scale challenges of the cold dark matter scenario. I tackle these issues in my PhD thesis by using the heterogeneous distribution of matter in the universe. I introduced the mechanism of cosmic structure formation with small primordial *perturbations* that began to grow when the universe became matter dominated. Dark matter grows more and more clumpy because of the attractive nature of gravity. An overdensity of 10^{-5} when $T \sim 1$ eV grew to 10^{-4} by the time the temperature dropped to 0.1 eV. Eventually, perturbations ceased to be small and became the *non-linear structures* we observe today. To understand these structures we must allow deviations from smoothness. We use statistical tools to study these perturbations around the smooth background that I introduce in Sec. 1.3.1.1 and I describe how it can be used to quantify the clustering of large-scale structures. In Sec. 1.3.1.2 I review the dynamics of gravitational instabilities that give rise to the observed structures. As long as density fluctuations are small enough, non-linearities can be treated by adopting perturbative expansion around linear solutions, but this breaks down on small scales and perturbation theory is non longer valid. While the impact of massive neutrinos and warm dark matter is the most significant on non-linear scales, I will introduce their effects on the matter clustering using linear theory in Sec. 1.3.2 and Sec. 1.3.3, which is a good approximation for a broad overview. However, I will use numerical theoretical predictions, which are far more accurate on small scales, in order to make precise constraints in the following Chapters.

1.3.1 Density fluctuations of large-scale structures

1.3.1.1 Statistical tools to probe inhomogeneities of cosmic fields

I introduced in Sec. 1.2.3 how cosmic structure grew from primordial inhomogeneities through gravitational instabilities. Early universe models predict the emergence of stochastic primordial fields coming from quantum fluctuations of the inflaton field. These primordial fluctuations are not observable directly, we only have access to the large-scale structures that we can study through the properties of their spatial distribution. Inflationary models predict the statistical properties of primordial fluctuations and their dynamical evolution depends on a cosmological model, i.e. the matter-energy budget, the theory of gravity and the expansion of the universe. Also, we only observe specific snapshots along lightcones. Therefore, in order to constrain cosmology using large-scale structures we need to adopt a statistical approach and to resort to a probabilistic description. All scalar cosmic fields, such as the density field or the gravitational potential field, are described as classical stochastic fields, meaning that at each point $\mathbf{x}$, $\delta(\mathbf{x})$ is a random variable. The universe is modeled as one specific stochastic *realisation* of a *statistical ensemble* of possible realisations.

In the following, we will consider the density perturbation field, or density contrast, a scalar field defined by

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \langle\rho\rangle}{\langle\rho\rangle}, \tag{1.41}$$

where we drop the time dependence and <> stands for ensemble average. However, we have no evidence that the ensemble exists since we are only able to observe one realisation. In cosmology, we rely on the common axiom assuming the ergodicity property of cosmic fields, meaning that the ensemble averages can be interpreted as volume averages if the considered volume is large enough. Therefore, the density contrast can be rewritten

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \overline{\rho}}{\overline{\rho}}. \tag{1.42}$$

There are strong observational evidences, in particular inflation, to assume that the primordial fluctuations that gave rise to the large-scale structures follow a Gaussian distribution. In addition, without resorting to inflation, from the central limit theorem, Gaussianity results from a superposition of a large number of random processes. If the amplitude of fluctuations is small enough this is a fairly good approximation, which appears to be the case by looking at CMB anisotropies that exhibit fluctuations around the smooth background at the level of 10^{-5}. An important property of Gaussian random fields is that the phases of the Fourier modes are randomly distributed. Therefore all the statistical information about the density perturbation field is contained in the variance of its amplitudes (its mean being $\langle\delta\rangle = 0$ by statistical homogeneity), i.e. the **two-point correlation function** (2PCF) or the **power spectrum**.

The 2PCF is defined as

$$\xi(r) = \langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r}) \rangle \,, \tag{1.43}$$

which only depends on $r = |\mathbf{r}|$ because of statistical homogeneity and isotropy. The physical interpretation is the measure of the excess over random probability that two objects are separated by the distance r. The Fourier transform of a Gaussian random field is also Gaussian and we can write it as

$$\tilde{\delta}(\mathbf{k}) = \int \delta(\mathbf{x}) e^{-i\mathbf{k}.\mathbf{x}} \, \mathrm{d}^3\mathbf{x}, \tag{1.44}$$

and equivalently

$$\delta(\mathbf{x}) = \int \tilde{\delta}(\mathbf{x}) e^{i\mathbf{k}.\mathbf{x}} \, \mathrm{d}^3\mathbf{x}. \tag{1.45}$$

$\tilde{\delta}(\mathbf{k})$ is the amplitude of the Fourier mode and $\mathbf{k}.\mathbf{x}$ its phase. In Fourier space, the correlation power spectrum is defined as

$$\left\langle \tilde{\delta}(\mathbf{k})\tilde{\delta}(\mathbf{k}') \right\rangle = (2\pi)^3 \delta_{\mathrm{dirac}}(\mathbf{k} - \mathbf{k}') P(k), \tag{1.46}$$

where δ_{dirac} is the Dirac distribution. Therefore, 2PCF and power spectrum are simply Fourier pairs and we can write

$$\xi(r) = \int P(k) e^{i\mathbf{k}.\mathbf{x}} \, \mathrm{d}^3\mathbf{x}. \tag{1.47}$$

If the auto-correlation function describes the probability of finding two objects at a distance r, the power spectrum decomposes this probability into characteristic lengths, $k = 2\pi/r$, and its amplitude describes the degree to which each characteristic length contributes to the total probability.

1.3.1.2 Evolution of density fluctuation in the linear regime

Given initial conditions for each mode $\delta(k)$ given by inflation, we want to find the model-dependent evolution equation of density fluctuations to compare their prediction with the matter power spectrum observed today and infer cosmological parameters. However, the time evolution of large-scale structures requires knowing how a small density fluctuations, $|\delta| \ll 1$, grows in amplitude under the influence of gravity. The study of cosmological perturbations can be fully treated in linearized GR, using the Einstein Field Equations. However, considering that perturbations are small, it implies that gravitational fields are weak so that much of the essential physics can be extracted adopting a Newtonian approach. Also, because growth of structures are driven by dark matter that dominates the matter energy budget I will only go through the evolution of dark matter density perturbations. At late times, well after the recombination, baryons are fully decoupled and follow dark matter perturbation such that we can consider that $\delta_b \sim \delta_{\mathrm{cdm}}$ deep in the matter-domination period while $\delta_b \sim \delta_{\mathrm{r}}$ well before recombination.

We consider a non-relativistic perturbated fluid in an expanding universe with density $\rho_{cdm} = \bar{\rho}_{cdm}(1+\delta_{cdm})$ for dark matter and $\rho_b = \bar{\rho}_b(1+\delta_b)$ for baryons, pressure $p = p_0 + \delta p$ and velocity $\mathbf{v} = \mathbf{v}_0 + \mathbf{u} = H\mathbf{x} + \mathbf{u}$ where $\mathbf{x}$ is the physical coordinate related to the comoving coordinate $\mathbf{q}$ by $\mathbf{x} = a\mathbf{q}$. The gravitational field is written as $\phi = \phi_0 + \delta\phi$. Considering that all perturbations are small, the fundamental equations governing its motion in comoving coordinates are

$$\begin{aligned} &\text{Euler equation} && \dot{\delta}_{cdm} + \tfrac{1}{a}\nabla\cdot\mathbf{u} = 0 \\ &\text{Continuity equation} && \dot{\mathbf{u}} + \tfrac{\dot{a}}{a}\mathbf{u} + \tfrac{c_s^2}{a}\nabla\delta_{cdm} + \tfrac{1}{a}\nabla\phi = 0 \\ &\text{Poisson equation} && \nabla^2\phi = 4\pi G a^2\bar{\rho}_m\delta_{cdm} \end{aligned} \tag{1.48}$$

with $c_s^2 = \frac{\partial p}{\partial\rho_m}$ is the sound speed acting as the equation of state closing the set of equation. For the Poisson equation, I used the fact that

$$\bar{\rho}_{cdm}\delta_{cdm} + \bar{\rho}_b\delta_b = (\bar{\rho}_{cdm} + \bar{\rho}_b)\delta_{cdm} = \bar{\rho}_m\delta_{cdm} \tag{1.49}$$

in the matter-domination period. If we take the divergence of the continuity equation to inject in the Poisson equation using also the time derivative of the Euler equation, we obtain a single, second-order differential equation

$$\ddot{\delta}_{cdm} + 2\frac{\dot{a}}{a}\dot{\delta}_{cdm} - \frac{c_s^2}{a^2}\nabla^2\delta - 4\pi G\bar{\rho}_m\delta_{cdm} = 0. \tag{1.50}$$

By Fourier transforming δ_{cdm} and applying the transformation $\nabla^2 \to -k^2$ we get

$$\ddot{\tilde{\delta}}_{cdm} + 2\frac{\dot{a}}{a}\dot{\tilde{\delta}}_{cdm} + c_s^2\left[\frac{k^2}{a^2} - \frac{4\pi G\bar{\rho}a^2}{c_s^2}\right]\tilde{\delta}_{cdm} = 0. \tag{1.51}$$

In the non-expanding case, we would have $a = 1$ and $\dot{a} = 0$ so the solution would simply be

$$\tilde{\delta}_{cdm}(t) = e^{\pm t/\tau}, \tag{1.52}$$

with

$$\tau = 1/\sqrt{4\pi G\bar{\rho} - c_s^2 k^2}. \tag{1.53}$$

We can define the physical Jeans wavenumber

$$k_J = \frac{\sqrt{4\pi G\bar{\rho}}}{c_s^2}, \tag{1.54}$$

and the associated physical **Jeans length** $\lambda_J = 2\pi/k_J$, which is the ratio between the gravitational dynamic time and the time scale of propagating pressure wave. We can distinguish two behaviors :

- if $k > k_J$ perturbations oscillate as sound waves.
- if $k < k_J$ perturbations are exponentially growing or decaying.

Indeed on small scales pressure is strong and inhomogeneities do not grow. This gravitational instability is the motor of large-scale structure formation. Also, the order in which scales undergo Jeans collapse will enable to distinguish between the bottom-up and top-down structure formation scenario. In the former, smaller scales undergo Jeans collapse first before merging into larger structures, while in the latter, larger scales collapse first before breaking up into smaller structures.

For an expanding universe analytical solutions are more complex but their characteristics are the same. For small scales $k \gg k_J$ we have oscillatory solution with frequency $w \sim c_s k/a$. To study perturbations in the large-scale limits we take $k \to 0$ and we end up with the following equation

$$\ddot{\tilde{\delta}}_{cdm} + 2\frac{\dot{a}}{a}\dot{\tilde{\delta}}_{cdm} - 4\pi G\bar{\rho_m}a^2\tilde{\delta}_{cdm} = 0 \tag{1.55}$$

also giving one decaying mode, D_- and one growing mode, D_+. Because the latter is responsible for structure formation, its analytical expression is called the **growth function**. Using the Friedmann equation for the expansion, we have in the matter-domination period $\bar{\rho}_{cdm} \propto a^{-3}$. We can then derive the cold dark matter perturbation growing mode,

$$\tilde{\delta}_{cdm} \propto D_+[\mathrm{MD}] \propto a. \tag{1.56}$$

In the radiation-domination period the Poisson equation reads

$$\nabla^2\phi = 4\pi Ga^2(\bar{\rho}_{cdm}\delta_{cdm} + \bar{\rho}_b\delta_b + \bar{\rho}_r\delta_r) \approx 0 \tag{1.57}$$

because $\bar{\rho}_{cdm}, \bar{\rho}_b \ll \bar{\rho}_r$ and $\delta_r \to 0$ given that pressureless master does not cluster. Hence the differential equation becomes

$$\ddot{\tilde{\delta}}_{cdm} + 2\frac{\dot{a}}{a}\dot{\tilde{\delta}}_{cdm} = 0. \tag{1.58}$$

Therefore, the growing mode for dark matter perturbations in the radiation-domination period is no longer linear with the scale factor but logarithmic,

$$\tilde{\delta}_{cdm} \propto D_+[\mathrm{RD}] \propto \ln(a). \tag{1.59}$$

Also, during the radiation domination period the sound speed is similar to the light speed. Thus the Jeans length is comparable to the horizon scale and perturbations inside the horizon do not grow. Thus, at early time, the dominant energy of radiation drives the universe to expand so fast that the matter has no time to respond and self-gravitates. Therefore, the perturbations inside the horizon are frozen at a constant value, while super-Hubble scales grow as $\ln(a)$.

1.3.1.3 The matter power spectrum

Under linear evolution, each mode $\tilde{\delta}(k,t)$ evolves independently and thus the evolved density is a linear functional of the initial conditions. Early universe models predict a nearly-scale invariant power spectrum parametrized as follow,

$$P_{\mathrm{init}}(k) = A_s\left(\frac{k}{k_0}\right)^{n_s-1} \tag{1.60}$$

where A_s is the amplitude of the primordial perturbation, the scalar index n_s measures the deviation from scale invariance ($n_s = 1$) and k_0 is the pivot scale. Hence, we expect the processed matter power spectrum later at some redshift z to write as

$$P_m(k,z) = T_k^2(z)P_{\mathrm{init}}(k), \tag{1.61}$$

where T_k is the transfer function of matter of the mode k and gives the ratio of the late time amplitude of a mode to its initial value, such that

$$T(k,z) = \frac{\tilde{\delta}_m(k,z)}{\tilde{\delta}_m(k,z=z_i)D_+(z)}, \tag{1.62}$$

with z_i an initial redshift. It only depends on the growing mode solution since we expect it to dominate after sufficient evolution.

We need to combine information on all scales, inside and ouside the horizon, before and after matter radiation equality. A scale *enters the horizon* when its length is comparable to the Hubble length $1/H$, this happens when $k = aH$. After inflation, the Hubble scale is growing in comoving coordinates so new scales constantly enter the horizon. The transition from radiation to matter domination occurs at $1+z_{\mathrm{eq}} = \Omega_m/\Omega_r$, corresponding to a comoving wavenumber $k_{\mathrm{eq}} = a_{\mathrm{eq}}H_{\mathrm{eq}}$. Thus, scales larger than the Hubble scale at equality ($k < k_{\mathrm{eq}}$) entered the horizon during matter domination. They have always been growing like $\ln(a)$ during the radiation domination, and like a during the matter domination, independently of their scale. However, scales smaller than the Hubble scale at equality ($k > k_{\mathrm{eq}}$) entered the horizon during radiation domination. Thus, they underwent a deficit of growth while inside the horizon during radiation domination.

The present-day matter power spectrum is defined as

$$P_m(k,z=0) = \left\langle |\tilde{\delta}_m(k,z=0)|^2 \right\rangle, \tag{1.63}$$

with $\tilde{\delta}_m = \frac{\sum_i \bar{\rho}_i \tilde{\delta}_i}{\sum_i \bar{\rho}_i}$ where the sum runs over all non-relativistic components. In a minimal flat, ΛCDM model with massless neutrinos it can be described with six parameters ; the cosmological constant density Ω_Λ, the non-relativistic matter density $\omega_m = \Omega_m h^2$, the baryon density $\omega_b = \Omega_b h^2$, the primordial spectrum amplitude A_s and logarithmic slope n_s and the optical depth to reionization τ. These parameters control various physical effects modifying the observable shape of the matter power spectrum. For instance, the time of radiation-matter equality is fixed by the radiation and baryon densities. Since ω_r is fixed by the CMB temperature, modifying ω_b will impact the overall shape. If we lower ω_b we consequently delay the time of radiation-matter equality, thus more modes enter the horizon during radiation domination. Since sub-Hubble scale perturbations grow less efficiently during matter-domination compared to radiation domination, decreasing ω_b suppresses power on small scales compared to large scales. The total of six parameters have each specific effects so that it is in principle easy to measure them using a combination of CMB and large-scale structures, assuming a flat ΛCDM model. However it is quite difficult to measure accurately the power spectrum of the total non-relativistic matter since it is essentialy made up of dark matter, whose presence can only be inferred indirectly. Thus, we measure the distribution of baryonic *tracers*, i.e. massive objects such as galaxies or quasars, which are *biased* tracers of the total matter distribution. We can relate the total matter power spectrum to the power spectrum of a biased tracer with

$$P_b(k,z) = b^2(k,z)P_m(k,z), \tag{1.64}$$

where $b(k,z)$ is the biasing function, which can be a function of many parameters depending on the considered tracer. Also, when measuring the power spectrum of a tracer, we must take into account redshift distorsion due to its peculiar velocity which adds a Doppler shift (blue-ward or red-ward) and skews the redshift along the line-of-sight. This anisotropy induced in the power spectrum is characterized by the kaiser β parameter (Kaiser, 1987),

$$P_b(k,z) = b^2(k,z)\left(1+\beta\mu^2\right)^2 P_m(k,z), \tag{1.65}$$

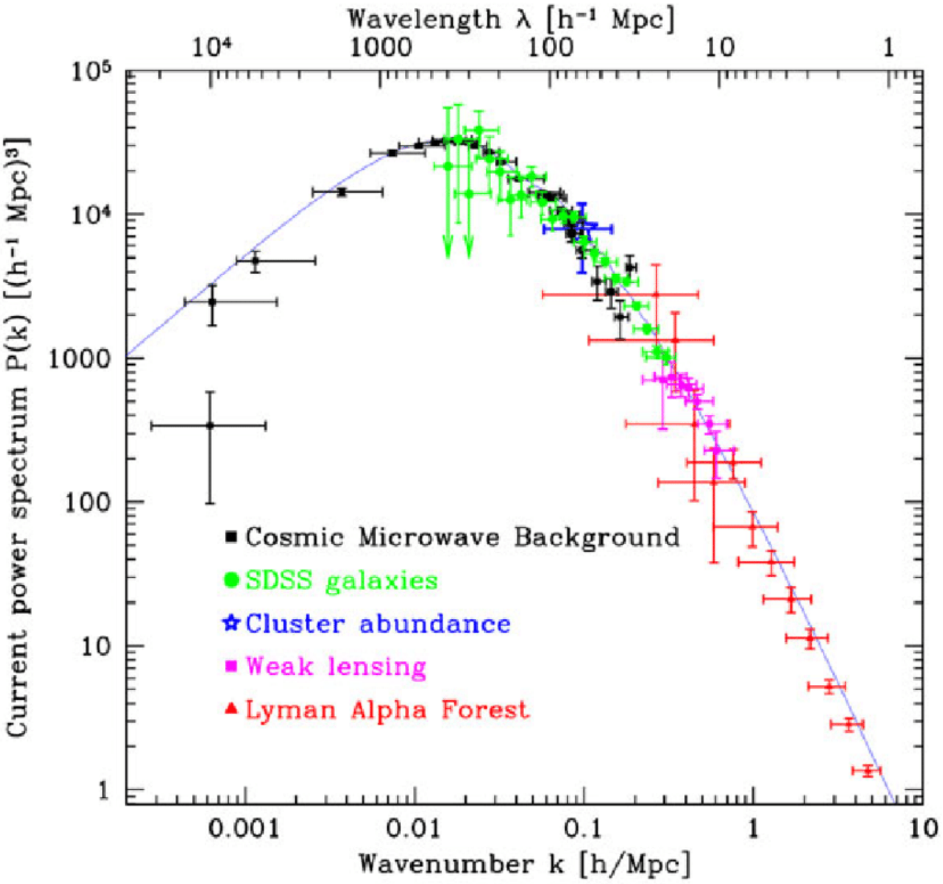

FIGURE 1.8 – *The infered matter power spectrum at $z = 0$ using CMB data on the largest scales, galaxy clustering and weak lensing on intermediates scales and the Lyα forest on the smallest scales. The red solid line is the linear prediction theory for a ΛCDM cosmological model. Tegmark and Zaldarriaga (2002). This plot is updated with the most up-to-date data in Chapter 7*

where μ is the cosine of the angle between the line of sight and the wavevector $\mathbf{k}$.

Fig. 1.8 shows the matter power spectrum today estimated from a combination of data covering a wide range of scales and epochs. The wavevector associated to the peak corresponds to the mode k_{eq} entering the horizon at the matter-radiation equality, as smaller scales are clearly suppressed. I will provide an update of this work using the most up-to-date data in Chapter 7.

To summarize, the matter power spectrum of large-scale structures is a key observable to constrain cosmology since it is efficiently parametrized by a set of 6 independent cosmological parameters. In the following section, I will describe how it is affected by the inclusion of massive neutrinos and the hypothesis of warm dark matter instead of cold dark matter.

1.3.2 Structures in massive neutrino model

In this part I will consider a simple extension of the ΛCDM model, the ΛCDMν model, adding the sum of the masses of the three neutrinos, $\sum m_\nu$, as an extra parameter. Given the range allowed by neutrino oscillation experiments from equation (1.40), we find that Ω_ν is restricted to the approximate range

$$0.0013 \leq \Omega_\nu \leq 0.13. \tag{1.66}$$

The neutrino density is fixed by the effective mass parameter $m_\nu^{\rm eff} = \sum m_\nu$. A $m_\nu^{\rm eff} = 93.14$ eV neutrino would have a critical energy density so we can write

$$\omega_\nu = \Omega_\nu h^2 = \frac{m_\nu^{\rm eff}}{93.14 \text{ eV}}. \tag{1.67}$$

Thus, even if neutrinos have the appropriate properties to make dark matter they are not heavy enough to account for the full lacking mass. Also, we will see that dark matter with a large velocity dispersion similar to that of neutrinos, called hot dark matter, erases the density fluctuations on small scales. A universe dominated by hot dark matter is in strong contradiction with various observations. For instance, large objects such as supercluster of galaxies would form first and smaller objects would form via fragmentation process. The obvious failure of a hot dark matter model justifies the consideration of the ΛCDMν model, where neutrinos act as hot dark matter component at late time in addition of a dominating cold dark matter component.

1.3.2.1 Neutrino free-streaming

After thermal decoupling neutrinos constitute a collisionless fluid, where the individual particles *free-stream* with a characteristic velocity that is on average the thermal velocity v_{th}. By analogy with the Jeans length, the radiation population can free-stream out of a characteristic horizon, which is the typical distance that they can travel without colliding with another particle. This collisionless damping, or **Landau damping**, is characterized by the free-streaming wavenumber defined as

$$k_{FS}(t) = \sqrt{\frac{4\pi G\bar{\rho}(t)a^2(t)}{v_{th}^2(t)}}, \tag{1.68}$$

and the free-streaming length as

$$\lambda_{FS}(t) = 2\pi \frac{a(t)}{k_{FS}(t)} = 2\pi\sqrt{\frac{2}{3}}\frac{v_{th}(t)}{H(t)}. \tag{1.69}$$

As long as neutrinos are relativistic they travel at the speed of the light and their free-streaming length is simply equal to the Hubble length. When they become non relativistic, their thermal velocity decays as

$$v_{th} \sim \frac{<p>}{m} \approx \frac{3.15 T_\nu}{m} \propto \frac{1+z}{m}. \tag{1.70}$$

During matter domination we thus have

$$\lambda_{FS}(t) = 8\frac{1+z}{\sqrt{\Omega_{m,0}(1+z)^3}}\left(\frac{1 \text{ eV}}{m}\right) \text{ h}^{-1}\text{Mpc}. \tag{1.71}$$

So the comoving free-streaming length is

$$\frac{\lambda_{FS}(t)}{a} \propto (1+z)^{1/2}. \tag{1.72}$$

Therefore, the comoving free-streaming length actually decreases during matter domination. For neutrinos becoming non-relativistic during matter domination, the comoving free-streaming length passes through a maximum $\lambda_{\rm nr}$, and a minimum $k_{\rm nr}$ for the comoving free-streaming wavenumber. This minimum wavenumber is found to be

$$k_{\rm nr} \approx 0.018\ \Omega_m^{1/2}\left(\frac{m}{1 \text{ eV}}\right)^{1/2} \text{ h Mpc}^{-1}. \tag{1.73}$$

This free-streaming effect prevents neutrinos to be confined into regions smaller than the free-streaming length since their velocity is large enough to escape from gravitational potential wells. Therefore, neutrino density fluctuations are suppressed on small scales with $k > k_{\rm nr}$. On scales large enough, i.e. modes with $k \ll k_{\rm nr}$, neutrino velocity is vanishing so perturbations are never affected by free-streaming and they behave like pure cold dark matter perturbations.

1.3.2.2 The matter power spectrum in ΛCDMν

We know that the non-relativistic transition of a least two neutrino species takes place during matter domination, thus we consider the matter power spectrum deep in the matter domination period. Therefore, the matter density does not only include baryons and cold dark matter but also neutrino density, so that $\Omega_m = \Omega_{cdm} + \Omega_b + \Omega_\nu$. I introduce the neutrino abundance parameter

$$f_\nu = \frac{\Omega_\nu}{\Omega_m}. \tag{1.74}$$

On large scales, with $k \ll k_{nr}$, we have seen that neutrino perturbations behave just like cold dark matter perturbations. If the neutrino masses are included, with Ω_m fixed, then the large-scale matter power spectrum is indistiguishable between the ΛCDM and the ΛCDMν models.

On small scales, with $k \gg k_{nr}$, the matter power spectrum of ΛCDMν is suppressed compared to that of ΛCDM because massive neutrino free-streaming prevents neutrino perturbations to grow, delays matter-radiation equality and finally slows the growth of cold dark matter perturbations at late time.

First, because massive neutrinos free-stream on small scales we have $\delta_\nu \to 0$ for $k \gg k_{NR}$. The matter power spectrum can be expanded as a function of the three non-relativistic species

$$\begin{aligned} P_m(k,z) &= \left\langle \left| \tfrac{\delta\rho_{cdm}+\delta\rho_b+\delta\rho_\nu}{\bar\rho_{cdm}+\bar\rho_b+\bar\rho_\nu} \right|^2 \right\rangle \\ &= \left\langle \left| \tfrac{\bar\rho_{cdm}\delta_{cdm}+\bar\rho_b\delta_b+\bar\rho_\nu\delta_\nu}{\bar\rho_{cdm}+\bar\rho_b+\bar\rho_\nu} \right|^2 \right\rangle \\ &= \left\langle \left| \tfrac{\Omega_{cdm}\delta_{cdm}+\Omega_b\delta_b}{\Omega_{cdm}+\Omega_b+\Omega_\nu} \right|^2 \right\rangle \\ &= \left(\tfrac{\Omega_{cdm}+\Omega_b}{\Omega_m}\right) \left\langle |\delta_{cdm}|^2 \right\rangle \\ &= (1-f_\nu)^2 \left\langle |\delta_{cdm}|^2 \right\rangle, \end{aligned} \tag{1.75}$$

where I used the fact that deep into the matter-domination period $\delta_b = \delta_{cdm}$. Thus, even if the growth of perturbations of cold dark matter were not affected by massive neutrinos, the power spectrum would be reduced on small scales because neutrinos do not cluster while accounting for non-relativistic matter.

Then, massive neutrinos reduce the growth rate of cold dark matter (and baryon) perturbations, first by delaying the matter-radiation equality. Indeed, we know that the non-relativistic transition of a least two neutrino species takes place during matter domination. Since we always have

$$\frac{a_{\rm eq}}{a_0} = \frac{\omega_r}{\omega_b + \omega_{cdm}} = \frac{1}{1-f_\nu}\frac{\omega_r}{\omega_m}. \tag{1.76}$$

It implies

$$\frac{a_{\rm eq}^{f_\nu \neq 0}}{a_{\rm eq}^{f_\nu = 0}} = \frac{1}{1-f_\nu}. \tag{1.77}$$

meaning that the ΛCDMν model is expanding faster. More generally, at any time before the non-relativistic transition, matter perturbations grow the same ways in the two models with a shift in the scale factor. So that for $a \leq a_{nr}$

$$\delta_{cdm}^{f_\nu \neq 0}(k,a) = \delta_{cdm}^{f_\nu = 0}\left(k,(1-f_\nu)a\right). \tag{1.78}$$

But more importantly, growth rate of matter perturbations are reduced because of an absence of gravitational back-reaction effects from free-streaming neutrinos. I recall that the evolution of non-relativistic matter in the linear regime is given by the combination of the continuity equation and the euler equation yielding

$$\ddot{\tilde{\delta}}_{cdm} + 2\frac{\dot{a}}{a}\dot{\tilde{\delta}}_{cdm} = -k^2\phi. \tag{1.79}$$

The right-hand side represents the gravitational clustering that is given by the Poisson equation as a function of the total matter density fluctuation. The left-hand side represents Hubble friction, meaning the enhancement of distances because of cosmological expansion that reduces gravitational attraction. In the absence of massive neutrinos the Poisson equation gives us

$$-k^2\phi = 4\pi G a^2(\bar{\rho}_{cdm}\tilde{\delta}_{cdm} + \bar{\rho}_b\tilde{\delta}_b) = 4\pi G a^2\bar{\rho}_m\tilde{\delta}_{cdm} \tag{1.80}$$

In the matter domination period the Friedmann equation gives $\rho_m \propto a^{-3}$, resulting to the growing mode solution $\tilde{\delta}_m \propto a$ when we inject it in equation (1.79). But in the case of free-streaming massive neutrinos, we neglect them from the Poisson equation since they do not cluster ($\delta_\nu \ll \delta_{cdm}$). However, they do participate to the homogeneous expansion through the Friedmann equation and we cannot neglect $\bar{\rho}_\nu$ compared to $\bar{\rho}_{cdm}$. Therefore, the exacte compensation between clustering and expansion in the ΛCDM model is slightly shifted in favour of the expansion effect in the ΛCDMν model. It is a pure background effect which leads to a modified evolution of matter perturbations. In presence of massive neutrinos, the equation driving the evolution of matter perturbation becomes

$$\begin{aligned} \ddot{\tilde{\delta}}_{cdm} + 2\tfrac{\dot{a}}{a}\dot{\tilde{\delta}}_{cdm} &= 4\pi G a^2(\bar{\rho}_{cdm} + \bar{\rho}_b)\tilde{\delta}_{cdm} \\ &= 4\pi G a^2(1-f_\nu)\bar{\rho}_m\tilde{\delta}_{cdm} \end{aligned} \tag{1.81}$$

Therefore, providing that $f_\nu \ll 1$ we can further derive that in the presence of massive neutrinos and for $k \gg k_{NR}$, the linear growth rate of matter perturbations is

$$\tilde{\delta}_{cdm} \propto a^{1-3/5 f_\nu}, \tag{1.82}$$

which is reduced compared to equation (1.56) because massive neutrinos contribute to the homogeneous expansion but not to gravitational clustering.

To summarize, the small-scale matter perturbation in presence of massive neutrinos is reduced because of the absence of neutrino perturbation in the total matter power spectrum and the slower growth rate of cold dark matter perturbations. The latter is the dominant effect. The total suppression can be approximated by the linear expression

$$\frac{P^{f_\nu \neq 0}}{P^{f_\nu = 0}} \approx -8 f_\nu. \tag{1.83}$$

Numerical simulations show that non-linear effects enhance the suppression that scales up to $\sim -10 f_\nu$. Fig. 1.9 shows this step-like suppression of the linear matter power spectrum, which

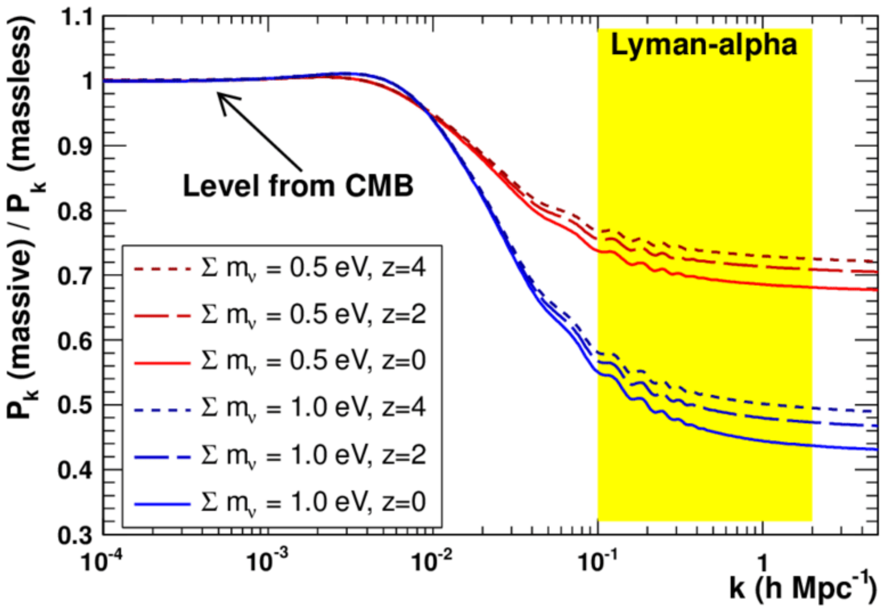

FIGURE 1.9 – *Linear theory prediction for the matter power spectra with massive neutrinos, normalized to the corresponding massless neutrino case for different masses and different redshifts. The yellow zone delimits the range of scales covered by the Lyman-α forest, the cosmological observable I use in the PhD thesis. From Palanque-Delabrouille et al. (2016)*

increases as $\sum m_\nu$ increases and as redshift decreases, leading to the conclusion that large-scale structures can put constraints on the maximal value of $\sum m_\nu$.

The suppression on the range $10^{-2}\ \mathrm{h \cdot Mpc^{-1}} \leq \mathrm{k} \leq 1\ \mathrm{h \cdot Mpc^{-1}}$ is particularly characteristic of massive neutrinos, and the redshift evolution makes it difficult to be mimicked by other cosmological parameters. However, no cosmological observable probes the full wavenumber range. For instance, galaxy clustering probes the transition region and the Lyman-alpha forest, which I will use in this thesis, probes the maximal suppression region. As such, massive neutrino effects are highly degenerated with any effects affecting the tilt of the power spectrum when using galaxy clustering, and any affects affecting the amplitude of the power spectrum when using the Lyman-alpha forest. It is common to lift these degeneracies using the CMB anisotropies that probe scales unaffected by the suppression.

Cosmology is mainly sensitive to $\sum m_\nu$ through their contribution to the energy density, i.e. through Ω_ν. We could go further and study the impact of the mass splitting between the different neutrinos. Because they do not become non-relativistic at the same time, the scale of the step-like suppression induced by each neutrino and the amount of suppression have a dependence of the individual masses. However, the changes are so small, at the level of 0.3% at most on the linear matter power spectrum, that even with very optimistic assumptions about future data, the mass splitting will remain undetectable (Archidiacono et al., 2020). However, it is important to stress that if cosmology could constrain $\sum m_\nu \leq 0.10$ eV it would rule out the inverted hierarchy scenario, which would be a major discovery. On the opposite, constraining $\sum m_\nu$ to be inferior to a value above 0.10 eV would not enable to make any conclusion on the mass hierarchy. Therefore, this PhD thesis uses cosmology to constrain the sum of the three neutrinos, $\sum m_\nu$, but does not probe the individual masses.

1.3.3 Structures in warm dark matter model

I presented in Sec. 1.2.4.2 the small-scale challenges faced by the cold dark matter scenario that predicts too many structures on scales below 10 $h^{-1} \cdot$ Mpc. Many theoretical models have emerged to solve these small-scale discrepancies. In the 1980's, hot dark matter constituted of neutrinos has been considered. At that time, it appeared as the most plausible dark matter candidate. However, such hot dark matter model implies a top-down structure formation and that the free-streaming of these relativistic particles destroy any fluctuations smaller than supercluster size, up to 40 Mpc for a eV neutrino, in strong disagreement with the observed distribution of galaxies. It has been given up since and only a portion of dark matter is assumed to be in the form of neutrinos.

A popular scenario among the many dark matter candidates are *superweakly* interacting particles (superWIMP), meaning their cross-section with standard particles is much smaller than that of WIMPS. This small strength of interaction implies that they were never in thermal equilibrium, or they decoupled very early, and that they were produced deep in the radiation domination period, while still being relativistic on the opposite of cold dark matter. Therefore, the density perturbations of these particles are suppressed below their free-streaming length. If the latter is comparable with galaxy scale, such a model is called a *warm* dark matter model. It has the advantage of interpolating the effects of cold dark matter on large scales with those of hot dark matter at sub-galactic scales. Therefore it is conveniently consistent with the distribution of large-scale structures, while circumventing the issues encountered by cold dark matter at small scales.

If such particles were initially in thermal equilibrium they decoupled well before neutrinos. Therefore their velocity dispersion is smaller and their free-streaming length is shorter than that of neutrinos. For instance, thermal relics with mass of the order of few keV have a free-streaming length comparable to galaxy scale, $\lambda_{FS} \sim 0.1$ $h^{-1} \cdot$ Mpc. The authors in Tremaine and Gunn (1979) put the critical Tremaine-Gunn bound at ~ 0.5 keV for fermionic dark matter to be considered as warm, since any lighter fermion is unviable dark matter candidate. As introduced in Sec. 1.3.2.1, when particles are relativistic enough, they free-stream to a scale λ_{FS} effectively unaffected by gravitational potentials. When assuming the total dark matter is made up of this warm particle, the matter power spectrum is suppressed below the free-streaming scale, λ_{FS} which is given by analogy with the neutrino in equation (1.69). I show in Fig. 1.10 the suppression of small-scale structures in simulations implementing warm and hot dark matter scenarios.

The transition between the relativistic regime and the non-relativistic regime takes place during the radiation domination era, and the velocity dispersion is given by $\sim c$ and by $\sim < p > /m$ respectively. We thus need the explicit distribution function $f(q)$ that might differ from a Fermi-Dirac (thermal) distribution depending on the production mechanism. The lower mass bound infered from the maximal free-streaming length (or equivalently minimal free-streaming wavenumber) depends on the production mechanism.

1.3.3.1 Warm dark matter candidates

Among viable warm dark matter candidates are **early-decoupled thermal relics**, which are thermalized particles of masses m_x of a few keV, that decouple deep within the radiation dominated era while relativistic (when $T \gg 100$ GeV). It includes gravitinos (Pagels and Primack, 1982), neutralinos (Lee and Weinberg, 1977), Marojon (Lattanzi and Valle, 2007) and other

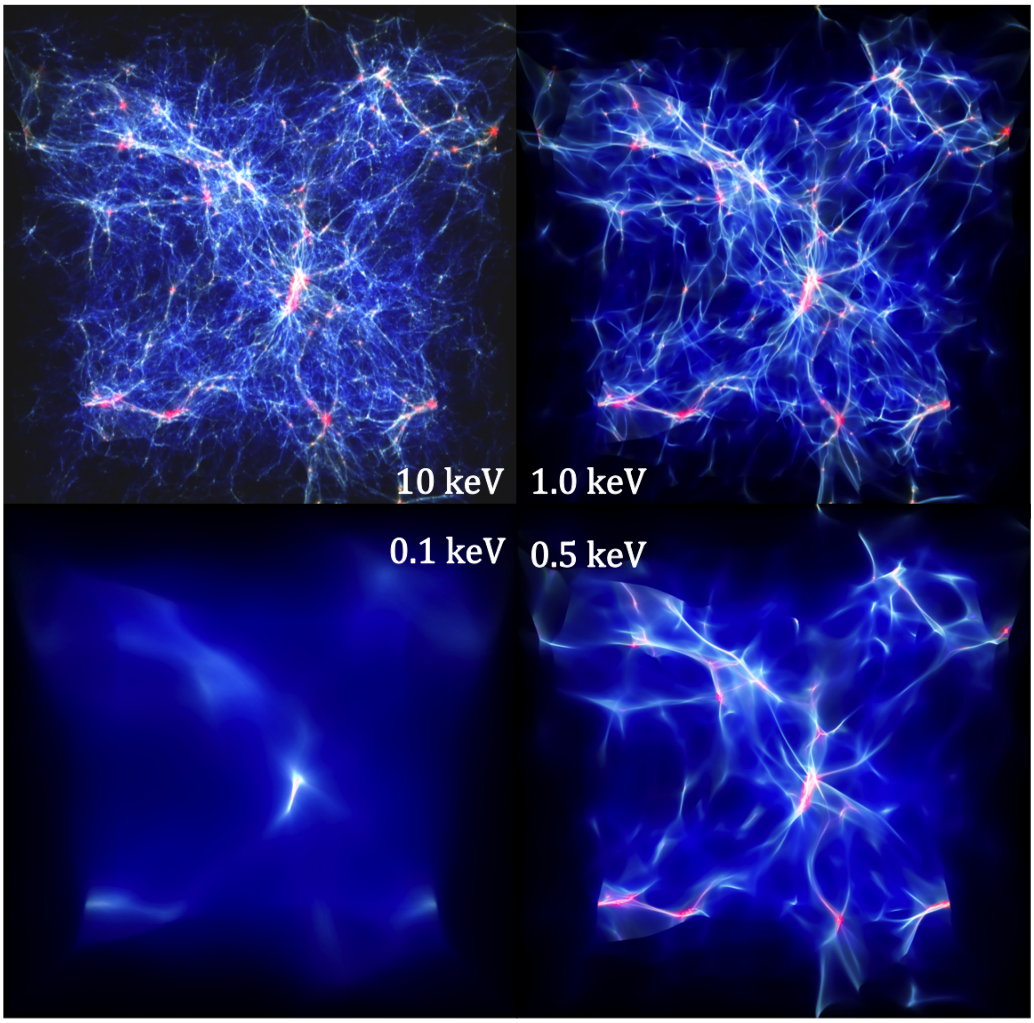

FIGURE 1.10 – *Distribution of gas particles in a (25 h.Mpc^{-1})^3 volume at $z = 2.5$ with different masses for warm dark matter particles with density and temperature encoded in intensity and color respectively. The two top panels are considered as warm dark matter, with the left one being indistinguishable from cold dark matter. The two bottom panels are considered as hot dark matter. From Baur et al. (2016).*

WIMPS. They are generally considered to have a momentum distribution following Fermi-Dirac statistics while in equilibrium (see equation (1.29)). However, the decoupling took place prior to neutrino decoupling. Therefore they do not share the entropy release from the different early-universe annihilations and their temperature is suppressed by a factor $T_x/T_\nu = \alpha < 1$. Also, any particle of temperature T_x coupled to photons prior to neutrino decoupling contribute as an additional ΔN_{eff} to the effective number of neutrinos, where

$$\Delta N_{\text{eff}} = \left(\frac{T_x}{T_\nu}\right)^4, \tag{1.84}$$

such that

$$N_{\text{eff}} = 3.046 \pm \Delta N_{\text{eff}}. \tag{1.85}$$

Finally, we can write the thermal relic energy density as

$$\omega_x = \left(\frac{T_x}{T_\nu}\right)^3 \frac{m_x}{93.14 \text{ eV}} = (\Delta N_{\text{eff}})^{3/4} \frac{m_x}{93.14 \text{ eV}}. \tag{1.86}$$

Another popular warm dark matter candidates are **sterile neutrinos**. Initially theorized by Asaka et al. (2005), sterile neutrinos, ν_s, are right-handed neutrinos (spin aligned with linear momentum, on the opposite of left-handed *active* neutrino), with no electric charge, no lepton charge and no color, making them insensitive to all fundamental interaction except gravity, and therefore very challenging to detect. There are no theoretical limitations to the scale of their mass, which can span from sub-eV to several 10^{15} GeV (Drewes, 2013). With one sterile neutrino of few keV, it would provide a theoretical framework to explain at the same time how active neutrinos acquire a mass, would account for CP violation and would be an adequate warm dark matter particle candidate.

Constraints on N_eff (see equation (1.33)) rule out $N_\nu = 4$ at the 5σ level. This means that if sterile neutrinos exist they must not have reached thermal equilibrium. There exist many production mechanisms occuring in the early universe. The simplest one is the Dodelson-Widrow mechanism (Dodelson and Widrow, 1994) that states that even if their interaction strength is too small to have been in thermal equilibirum with the primordial plasma they can be produced by non-null mixing with active neutrinos at $T \leq 1$ GeV. In the absence of lepton assymetry this production by oscillation reaches a maximum at $T_\mathrm{peak} \sim 150(m_{\nu_s}/\ \mathrm{keV})^{1/3}$. The resulting momentum distribution is approximately that of a reduced Fermi-Dirac, with its temperature equal to that of active neutrinos,

$$f_{\nu_s}(p) = \frac{\chi}{e^{p/T_\nu} + 1}, \tag{1.87}$$

The normalization factor $\chi \ll 1$ is determined by the mass of the sterile neutrino m_{ν_s} and dark matter abundance with

$$\omega_{\nu_s} = \chi \frac{m_{\nu_s}}{93.14\ \mathrm{eV}}. \tag{1.88}$$

This sterile neutrino production mechanism is also called *non-resonant production*. Since they share the same temprature than that of neutrinos, they contribute to an additional ΔN_eff to the effective number of neutrinos, with

$$\Delta N_\mathrm{eff} = \chi \left(\frac{T_\nu}{T_\nu}\right)^4 = \chi. \tag{1.89}$$

We can thus write the sterile neutrino energy density as

$$\omega_{\nu_s} = \Delta N_\mathrm{eff} \frac{m_{\nu_s}}{93.14\ \mathrm{eV}}. \tag{1.90}$$

In the presence of a net letpon asymmetry, sterile neutrinos are instead produced via resonant oscillations. Then the resulting sterile neutrinos have a cooler distribution with non-thermal features and smaller mixing angles. The transfer functions for resonantly-produced sterile neutrinos need to be specifically calculated with numerical simulations since non-thermal distributions significantly modify the evolution of perturbations in highly non-linear way (Horiuchi et al., 2015; Bozek et al., 2016). The Shi-Fuller mechanism (Shi and Fuller, 1999), the decay of a scalar field (Shaposhnikov and Tkachev, 2006; Kusenko, 2006; Petraki and Kusenko, 2008) or diluted thermal overproduction (Asaka et al., 2006; Bezrukov et al., 2010) belong to such class of sterile neutrino production mechanisms.

If enough keV sterile neutrinos are present in the universe, some of them will decay and produce a non-negligible amount of X-ray photons. Recent detections of an unidentified feature at 3.55 keV in the X-ray spectra of the Andromeda galaxy and the Perseus cluster were interpreted as a hint for a $m_{\nu_s} \sim 7.5$ eV non-resonantly produced sterile neutrino (Bulbul et al., 2014; Boyarsky et al., 2014; Cappelluti et al., 2017) but its origin is still highly debated (Merle and Schneider, 2015; Dessert et al., 2018; Boyarsky et al., 2020).

1.3.3.2 Matter power spectrum in ΛWDM

In this PhD work, I will only test the plausibility of a warm dark matter model formed entirely by early-decoupled thermal relics with Fermi-Dirac momentum distribution or by non-resonantly produced sterile neutrinos with a quasi-Fermi distribution normalized by $\chi \ll 1$. Given the similarities in their phase-space distribution function we can establish a mapping between m_x and m_{ν_s} (Colombi et al., 1996; Viel et al., 2005). If we consider the dark matter is entirely made up of thermal relics or sterile neutrinos, i.e. by setting $\omega_{DM} = \omega_x = \omega_{\nu_s}$, we can write

$$m_{\nu_s} = \kappa m_x^{\mu} \left(\frac{0.25 \times 0.7^2}{\omega_{dm}} \right)^{1/3} . \tag{1.91}$$

The authors in Viel et al. (2005) find $\kappa = 4.43$ keV and $\mu = 4/3$, while the authors in Bozek et al. (2016) find $\kappa = 3.90$ keV and $\mu = 1.294$ when accounting for additional physical effects in the non-resonantly produced sterile neutrinos. This relation offers a direct correspondance between m_x and m_{ν_s}, so that every impact of thermal relics on the matter power spectrum can be mimicked by a non-resonantly produced sterile neutrino with a mass following equation (1.91). Thus, in the following of this Chapter I will only go through the effects of thermal relics on the matter distribution.

By analogy with the transfer function giving the evolution of a mode k in the ΛCDM model, it is convenient to define the transfer function of the ΛWDM model that depends on the ratio of the matter power spectrum of the two models,

$$T^2(k,z) = \frac{P_{wdm}(k,z)}{P_{cdm}(k,z)}. \tag{1.92}$$

It is well fitted by the suggested formula by (Bode et al., 2001) :

$$T(k,z) = \left(1 + \left(\frac{k}{k_0} \right)^{2\nu} \right)^{-\nu/5} , \tag{1.93}$$

where ν is found to be 1.12 using Boltzmann code simulation in Viel et al. (2005) and k_0 is the cut-off scale. This describes a cut-off with k^{-10} dependence on scales $k \gg k_0$.

Unlike in the case of active neutrinos, the wavenumber at which P_{wdm} starts deviating from P_{cdm} is not the free-streaming wavenumber at the non-relativistic transition k_{FS}^{nr}, which is the minimal free-streaming wavenumber in the case of active neutrinos. Indeed, the cut-off scale is the free-streaming horizon, below which scales are effectively unaffected by gravitational potentials. It is defined as

$$\lambda_{FSH}(a) = \int_0^a \frac{< v >}{a' H^2} da'. \tag{1.94}$$

For active neutrinos, one can found that $k_{FS}^{nr} \sim (k_{FSH})_{\min}$. However, this is not the case for thermal relics. Since the non-relativistic transition takes place during the radiation domination period the free-streaming scale remains constant between a_{nr} and a_{eq}. Thus, it significantly differs from the free-streaming horizon. Indeed, for thermal relics the transition takes place during the radiation domination period. Given the distribution function of thermal relics one can derive the thermal velocity distribution and one finds that k_0 depends on the particle mass and cosmological parameters with

$$k_0 = 0.24 \left(\frac{m_x}{1 \text{ keV}} \right)^{-0.83} (\Delta N_{\text{eff}})^{0.83/4} \left(\frac{\omega_x}{0.25(0.7)^2} \right)^{-0.16} \text{Mpc}. \tag{1.95}$$

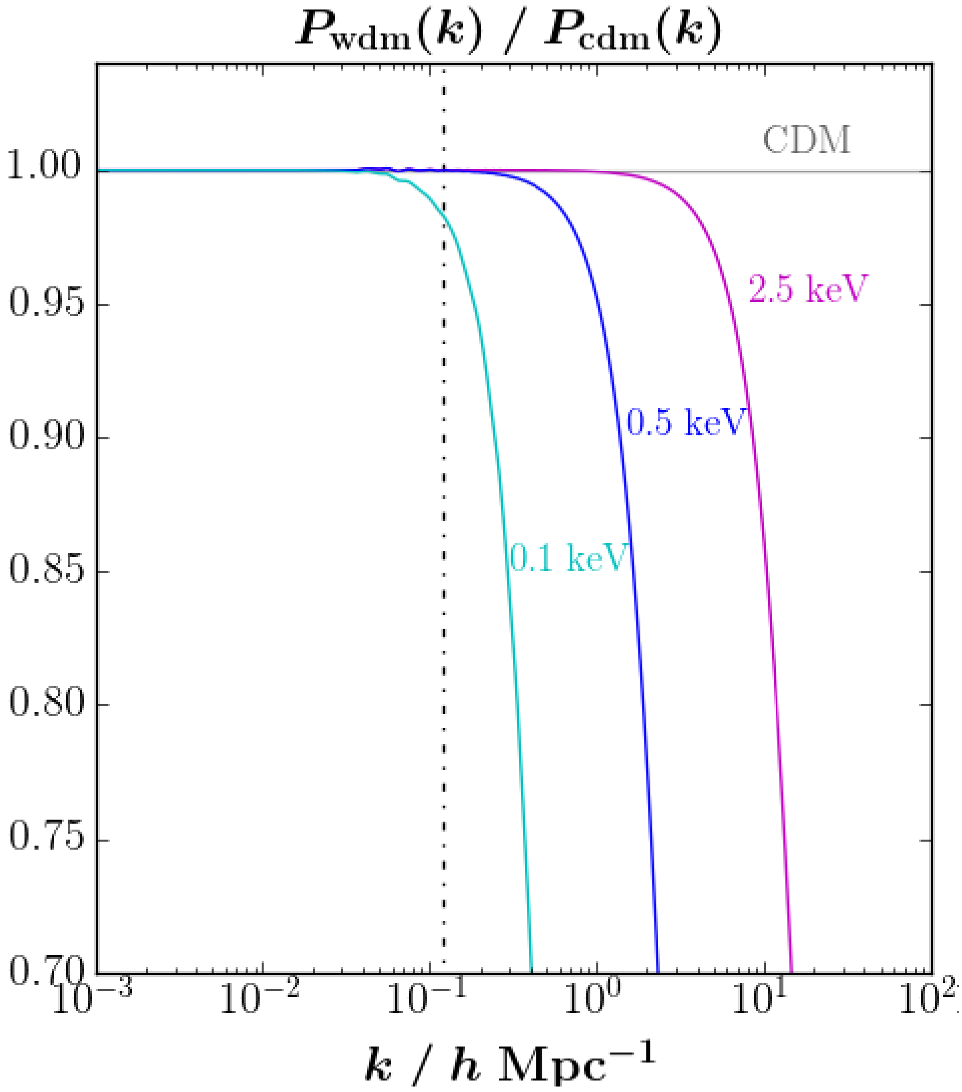

FIGURE 1.11 – *Analytical approximation of the linear matter power spectrum transfer function* $T^2(k, z = 0) = P_{wdm}(k, z = 0)/P_{cdm}(k, z = 0)$ *for 0.1 keV(cyan), 0.5 keV (blue) and 2.5 keV (magenta) dark matter particles. The region to the right of the dashed line indicates the range of scales probed by the Lyman-*α *forest, the cosmological observable in use in the PhD thesis. From Baur et al. (2016).*

I show in Fig. 1.11 the transfer functions for different thermal relic masses, where the warm dark matter power spectrum is indistinguishable from the cold dark matter one for low values of k (large scales). The cut-off scale materializes the free-streaming horizon scale which occurs at lower values of k as m_x decreases. The heavier (colder) the warm dark matter particle, the more compatible with cold dark matter its is. Therefore, the matter distribution on small scales is an efficient tool to probe the plausibility of a warm dark matter model since it prevents matter clustering below a characteristic scale that depends on its mass and offers a very characteristic signature.

1.4 Outline of this thesis

This brief overview of modern cosmology shows that there are still many roads of improvements regarding the standard model and many discoveries are still to be made. In particular, thanks to the disovery of the oscillations of neutrino flavours, we know that cosmological neutrinos are massive, whose influence spans the fields of particle physics and cosmology. Also, the cold dark matter paradigm appears to be in tension with observations on scales smaller than the Mpc. This motivates to consider a warm dark matter model, which is conveniently consistent with the distribution of large-scale structures, while circumventing the issues encountered by cold dark matter at small scales.

Massive neutrinos and warm dark matter models involve relativistic particles, i.e. active neutrinos, sterile neutrinos or thermal relics, whose free-streaming suppresses or alleviates formation of cosmic structures on small scales, below 1 Mpc. They produce characteristic signatures on the matter power spectrum, with a redshift evolution making it difficult to be mimicked by other cosmological parameters, as seen in Fig. 1.9 and in Fig. 1.11. Therefore, matter clustering on scales affected by relativistic free-streaming can put strong constraints on the mass of active neutrinos and on the mass of hypothetical warm dark matter particles. As such, cosmological surveys are ideally complementary to direct detection particle physics experiments, and the strong reduction of the allowed parameter space can efficiently guide the design of future experiments.

The Lyman-α (Lyα) forest, the collection of absorption lines in the spectra of distant quasars, is a powerful tool to study clustering in the universe at redshifts 2 to 6, on scales that are strongly non linear today, but were only midly non-linear at such high redshifts. The Lyα forest efficiently probes scales the most affected by the free-streaming of relativistic particles, as indicated in Fig.s 1.9 and 1.11 and extends the level arm of other cosmological probes. Also, the wide range of redshifts probed enable to test the redshift evolution of the models. As such, the Lyα forest has often be used to put the strongest bounds to date on $\sum m_\nu$ (Seljak et al., 2003; Viel et al., 2010; Palanque-Delabrouille et al., 2015,?; Yèche et al., 2017) and dark matter models suppressing small-scale cosmic structures such as warm dark matter (Narayanan et al., 2000; Viel et al., 2005; Seljak et al., 2006; Boyarsky et al., 2009; Viel et al., 2013; Baur et al., 2016, 2017; Garzilli et al., 2019), interacting dark matter (Dvorkin et al., 2014; Xu et al., 2018; Garny et al., 2018) or fuzzy dark matter (Armengaud et al., 2017; Nori et al., 2019; Iršič et al., 2017).

This justifies the use of the Lyα forest in this thesis, as the main cosmological probe to constrain the sum of the neutrino masses by putting an upper bound on $\sum m_\nu$, and to constrain warm dark matter models by putting a lower bound on the mass of hypothetical thermal relics, m_x and of hypothetical non-resonantly produced sterile neutrinos, m_{ν_s}.

In Chapter 2 I describe the formation of the Lyα forest produced by intervening neutral hydrogen clouds in the diffuse intergalactic medium. I will also show how it can be statistically used to constrain cosmology. In particular, I will introduce the 1D power spectrum of the Lyα forest that efficiently probes small scales along line of sights. Chapter 3 presents an overview of the Sloan Digital Sky Survey (SDSS) telescope with a brief description of the technical characteristics of the instruments and a special attention to the eBOSS spectroscopic survey. The SDSS is one of the largest, most ambitious and influential surveys in the history of astronomy, and I had the privilege to use an unprecedented high-quality 6-year data sample of high redshift quasars to conduct my PhD work. In Chapter 4, I present the measurement that I led of the 1D power spectrum of the Lyα forest of the BOSS and eBOSS programs. Constructing it from spectroscopic data requires a rigorous and detailed analysis and a thorough investigation of any potential observational systematics that could bias our scientific interpretation. Because the Lyα forest probes midly non-linear scales and because we observe a light flux, theoretical predictions have to be constructed with cosmological hydrodynamical simulations, whose concepts are outlined in Chapter 5. In Chapter 6 I present the Horizon-AGN series of hydrodynamical simulations, and the Extreme-Horizon simulation, which I performed using the Très Grand Centre de Calcul supercomputer to constrain the impact of Active Galactic Nuclei and stellar winds on the Lyα forest. Indeed, baryonic feedback strongly modify the thermal state and gas distribution in the intergalactic medium. As such, they are fully degenerate with cosmological parameters if not precisely accounted for. Chapter 7 compiles the results of the analysis that compares observations with predictive models. It provides the strongest constraints to date on the sum of the masses of active neutrinos, determines whether or not a warm dark matter model is allowed by observations and studies a weak persistent tension between the Lyα forest and CMB data. Finally, the outcome of this thesis is concluded and discussed in Chapter 8 along with prospects for future works.

2

The Lyman-α forest

Contents

In this work I use the Lyα forest as cosmological observable, which probes the range of scales where the suppression of matter density fluctuations caused by massive neutrinos and hypothetical warm dark matter particles is maximal. In Sec. 2.1 I introduce quasi-stellar objects that are the most luminous objects in the universe and serve as background light illuminating gas in the intergalactic medium. Studying absorption lines in their spectroscopic spectra provides valuable information from intervening materials along the lines of sight. In particular neutral clouds of hydrogen scatter light in the region blueward of the Lyα emission line and form the so-called Lyα forest. In Sec. 2.2 I introduce the physics of the Lyα forest and how it can be used in cosmology in Sec. 2.3.

2.1 Quasars

2.1.1 Active Galactic Nuclei

Active galaxies are galaxies with a small-emission core embedded at its center, which are capable of producing jet of gas 10^5 pc long. These central nuclei are called Active Galactic Nuclei (AGN). Much of the energy output is composed of non-thermal (does not come from stars) emissions over a broad range of frequencies, from radio to X-rays. Their luminosity can reach up to $10^{15}\,\mathrm{L}_\odot$ and they present a strong variability in time, as short as days, implying that the energy-emitting source must be of the order of light hours or light days in size. Their small size along with their prodigious energy emission led theorists to consider gravity as their energy source immediatly upon their discovery in the 1960's (Matthews, 1963). The gravitational potential energy can only be recovered if the AGNs contain compact objects with masses of $\sim 10^8 - 10^9\,\mathrm{M}_\odot$. Among the many potentiel candidates, the hypothesis that super-massive black holes (SMBH) of billions of solar masses power the AGNs is strongly supported by observational evidences. The radiated energy we detect comes from matter being accreted onto the black hole (BH) that forms an accretion disk surrounded by a dusty torus, in which some source of viscosity drains the orbiting matter from angular momentum, making it spiral inwards toward the central BH, as illustrated in Fig. 2.1.

The classification of AGNs is based on their special emission spectra in different wavelength ranges that cannot be explained by the superposition of stellar spectra. We can classify them upon their luminosity or radio-loudness :

- **Seyfert galaxies** are discovered in 1943 by the astronomer Carl Seyfert. They show an unsually bright point-like nuclei, with a variable excess of radiation in the far infrared compared to other galaxies. Their spectra exhibit broad lines with widths about $10,000\,\mathrm{km}\cdot\mathrm{s}^{-1}$ and narrow lines with widths up to $400\,\mathrm{km}\cdot\mathrm{s}^{-1}$.

- Quasi-stellar objects (QSO), or **quasars**, are also point-like objects but are the most luminous visible objects in the universe, much more luminous than Seyfert galaxies, with luminosities between 10 - 1,000 times the one of the Milky Way and only a hundred times the one of the Milky Way for Seyfert galaxies. A quasar is an extrem AGN in the sense that its intrinsic luminosity overcomes the luminosity of the host galaxy. Their spectra are similar to that of Seyfert galaxies but with weaker narrow-emission lines relative to broad lines. These objects serve as the background light for the matter tracer I use in this thesis, the Lyα forest, because they can be seen billions of light years away.

- **Blazars** are radio sources that do not display emission lines in their spectra and can vary significantly in brightness in less than a day, indicating that most of their light must come from a region less than one light-day across.

- **Radio-loud galaxies** are strong radio emitters with one or two radio lobes straddling the galaxy. The emission is higly non thermal and is due to the ejection of energetic particles, called outflows, which occur along the poles of the disk, escaping and forming collimated radio-emitting jets. These outflows interact with the interstellar and intergalactic medium, creating a plasma in which electrons spiral along magnetic-field lines and emit synchrotron radiation at radio wavelengths. Quasars and Seyfert galaxies can be radio-loud or radio-quiet as opposed to blazars that are always radio-loud.

The differences in these classes of AGNs are thought to arise from the orientation of the observation, and not necessary from different physical processes or different objects. It is nicely summarized in the orientation-based-unified model, as shown in Fig. 2.1. If the line of sight (LOS) is parallel to the accretion disk, the hot inner region is obscured by the dusty torus but the radio lobes, streching out to either one of both sides, will be clearly visible and the source is called a radio-loud galaxy. If the LOS is tilted compared to the accretion disk we will see both the blackbody from the inner regions of the disk and the synchrotron emission from the jets. Then, depending on its luminosity, it will be a radio-loud Seyfert galaxy or a radio-loud quasar. Finally, if the accretion disk is seen face-on, the jets are pointed at the Earth and appear extremely bright from radio to gamma rays, due to Doppler beaming. The resulting radio source is the blazar. Therefore in this model, the differences between AGNs arise simply because of the different orientations from which they are observed.

In the following, I will only focus on quasars that are the brightest source of light in the universe, such that they are the farthest objects that have been observed, the farthest one being at $z = 7.085$ (Mortlock et al., 2011). The extended Baryon Oscillation Sky Survey (eBOSS), described in Chapter 3, counts 659 quasars with $z > 5$. In addition to providing our first view of the universe beyond $z > 5$, quasars are also valuable cosmological probes. Indeed, they are luminous enough to act as background light to trace the matter distribution at very distant scales, which can be done by studying their spectra.

2.1.2 Quasar Spectrum

2.1.2.1 Continuum emission

Quasars emit radiations from γ-rays and x-rays to far-infrared wavelengths. On the contrary of thermal radiation from stars, which are peaked and restricted in wavelength, the emitted energy in each band is remarkably similar for every quasars. It is called the quasar continuum. Most quasars at $z < 2.5$ are bright at ultra-violet (UV) wavelengths. This is the key property to distinguish them from stars in sky surveys, which are more numerous but usually fainter in UV. The quasar-continuum emission appears to arise from a combination of thermal and non-thermal processes. In any event, the continuum radiation from quasars demonstrates that some very energetic processes are involved. Furthermore, the continuum radiation at the highest energies

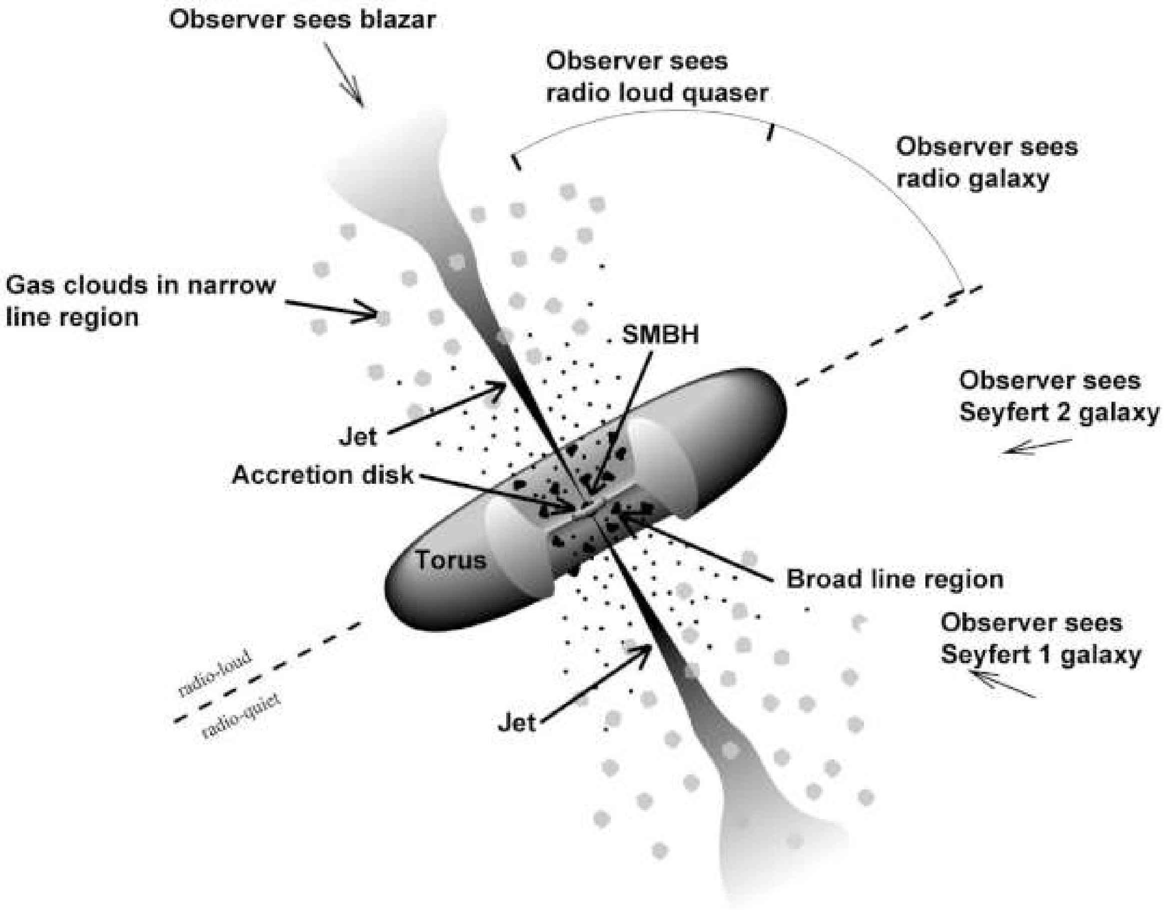

Figure 2.1 – *Schematic view of the dust-torus-based unified model of AGNs, where the orientation to which it is observed determines its class.Credit : Fermi collaboration.*

tends to show the most variability and the shortest timescales, which is another indication of the extreme conditions that exist near quasars.

The strongest emission lines in quasar spectra come from hydrogen, carbon and magnesium, with lines of nitrogen, oxygen, iron and other elements also being visible. The observed levels of ionization range from neutral for hydrogen and oxygen to five-times-ionized oxygen and even more highly ionized iron. Their width is determined by the gas velocity dipersion in the emitting region, where a mixture of infall, rotation and ejection probably occurs. The widths are consistent with the emission region being at a distance of light months to a few light-years from a central black hole. In Fig. 2.2 I show an eBOSS spectrum, with the estimated continuum radiation in red exhibiting multiple characteristic emission lines among H I, N V, C IV, He II and C III.

2.1.2.2 Absorption lines

Absorption lines in quasar spectra were discovered few years after their initial discovery and were initially believed to be quite rare. Nowadays, they represent a very active field of study in astrophysics and cosmology as they enable the study of gas along the photon trajectories. We distinguish three types of absorpting gas depending on their distance to the central engine source of the quasar :

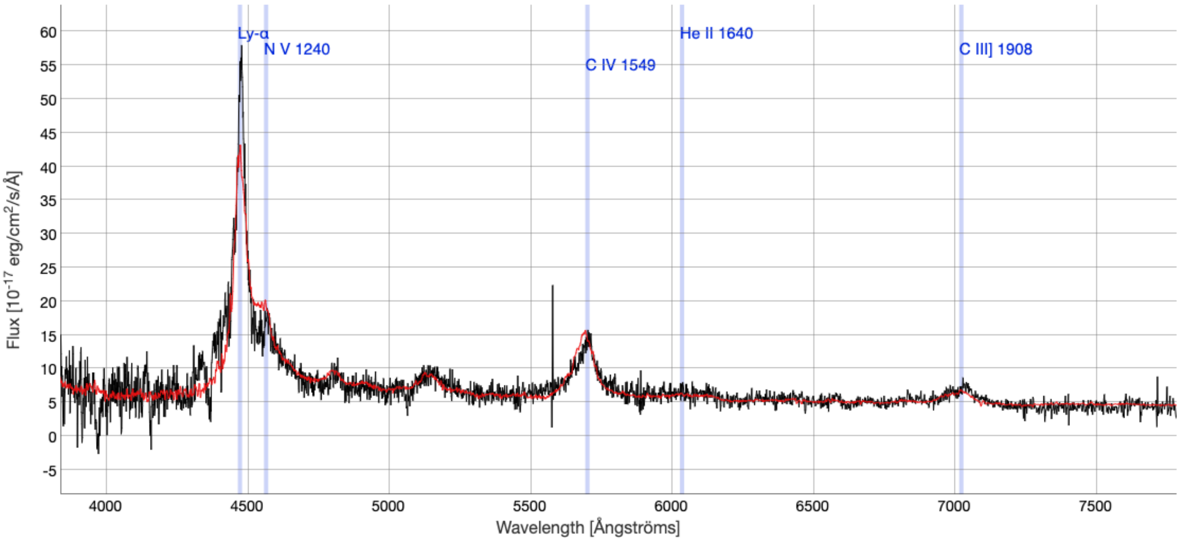

Figure 2.2 – *eBOSS quasar spectrum (SDSS J152626.60+412939.4) at redshift* $z = 2.68027$ *with the observed flux in black, the estimated continuum emission in red and the corresponding emission lines in blue.*

- Intrinsic systems arise from the quasar itself, either from the accretion disk or from the collimated jets straddling the central AGN. Broad Absorption Lines (BAL) are examples of such intrinsic systems. They appear blueward from the emission lines, with width indicating outflow velocities about $30,000\,\mathrm{km.s^{-1}}$. An example of a spectrum ehibiting BALs is shown in Fig. 2.3.

- Associated systems lead to narrower absorption lines, indicating lower velocity dispersion and have redshift close to the quasar one. Thus, they are thought to come from absorbing gas from the host galaxy or its close environment.

- Intervening systems have redshifts lower than the quasar-emission lines and arise in clouds of gas unrelated to the quasar that lie along the LOS. In this case, the quasar serves as a background beacon that enables the study of the gas that often do not emit any light. As such, the modification of the quasar continuum appears as the only direct observational evidence we have of this intervening gas.

Depending on the column density of the intervening gas, we can further divide them in three categories :

- **Damped Lyman-α Systems** (DLA), with column densities $\mathrm{N_{HI}} > 10^{20.3}$ H I atoms per cm^2, lead to very wide saturated absorptions in quasar spectra. They arise from gas clouds completely self-shielded against ionizing radiations from the outside, yet diffuse enough to have low star-formation rate. Such absorptions are thought to occur when the line of sight between the observer and the QSO comes close to a galaxy, throught its relatively dense circum-galactic medium (CGM), or even through the galactic disks in extreme cases. DLAs are a very active field of research as they give valuable information on the CGM of galaxies. Hence, DLAs can constrain galaxy formation and evolution or can

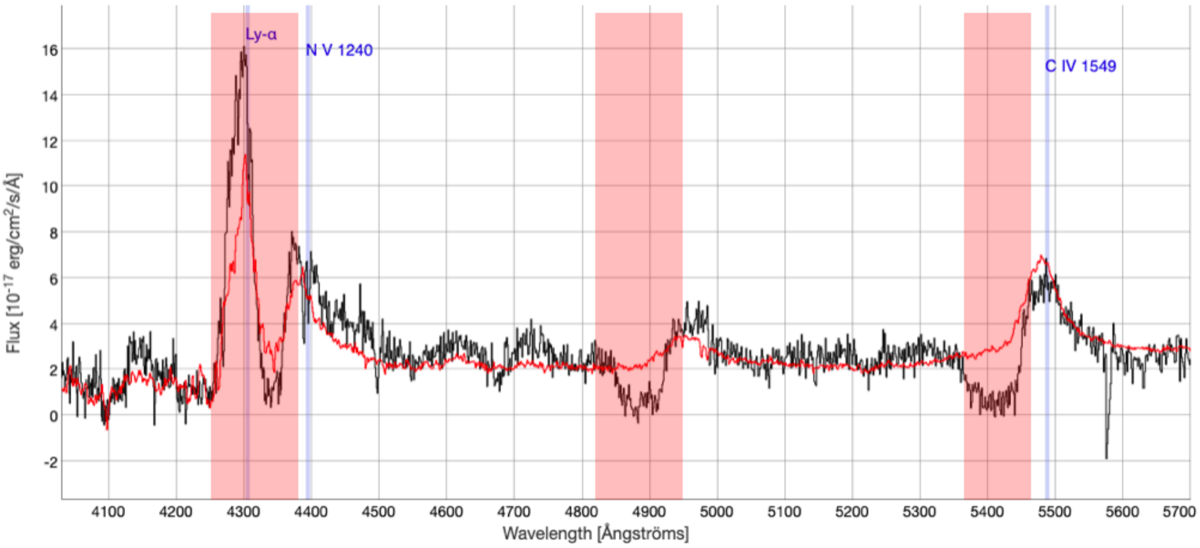

FIGURE 2.3 – *eBOSS quasar spectrum (SDSS J000001.76-072909.3) at redshift $z = 2.542323$ with the observed flux in black, the estimated continuum emission in red and the corresponding emission lines in blue. Broad absorption lines (red regions) are clearly visible redwards of the Lyα and C IV emission lines.*

act as tracers of matter. An example of a spectrum exhibiting a DLA is shown in Fig. 2.4.

- **Lyman-limit** or intermediate systems, with $10^{16} \leq N_{HI} \leq 10^{20.3}\text{cm}^{-2}$, arise where a gas cloud starts self-shielding itself. LLS are commonly attributed to the gas present in galaxy halos.

- The **Lyα forest**, with column densities $10^{12} \leq N_{HI} \leq 10^{16}$ cm^{-2}, arises from absorptions in the intergalactic medium (IGM). An example is shown in Fig. 2.5

The Lyα forest absorption lines are the most numerous ones in quasar spectra. More importantly, they are the only observational evidences we have of the IGM, which represents $\sim 90\%$ of the universe in volume : indeed, it does not emit any light hence it can only be seen in absorption.

2.2 The study of the IGM through the Lyα forest

2.2.1 The Lyα resonance absorption line

The Lyman series of hydrogen correspond to the set of UV-emission lines due to the electronic transition from an excited state $n \geq 2$ to the ground state $n = 1$, with n the principal quantum number. The Lyman-α (Lyα) and Lyman-β (Lyβ) are the first and second transitions of the Lyman series, from the ground $n = 1$ to the $n = 2$ and $n = 3$ states respectively. The Lyα transition corresponds to the absorbed (emitted) wavelength $\lambda_{\text{Ly}\alpha} = 1215.17$Å, and $\lambda_{\text{Ly}\beta} = 1026$Å for the Ly$\beta$ transition, in the rest-frame of the absorbers (emitters). Lyα and Lyβ photons have cross sections large enough for exciting a neutral hydrogen atom at its ground state. When the atom falls back to $n = 0$, the photon is not re-emitted in the same direction, it is

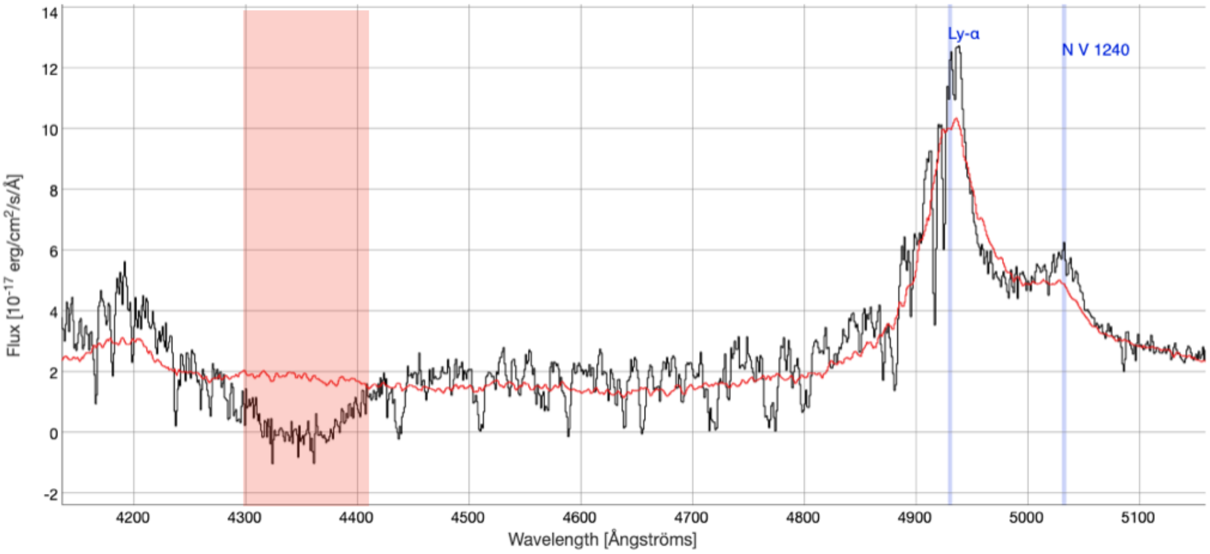

FIGURE 2.4 – *eBOSS quasar spectrum (SDSS J141501.31+532438.5) at redshift* $z = 3.056348$ *with the observed flux in black, the estimated continuum emission in red and the corresponding emission lines in blue. The saturated Damped Ly*α *absorption (red region) is clearly visible with the observed flux well below the quasar continuum.*

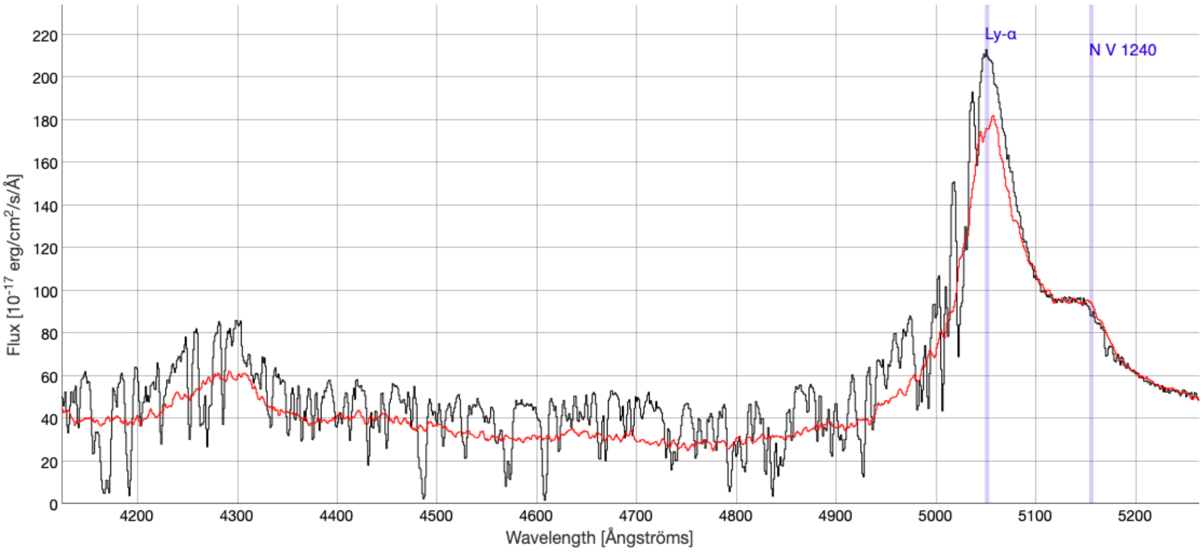

FIGURE 2.5 – *eBOSS quasar spectrum (SDSS J114308.87+345222.2) at redshift* $z = 3.155227$ *with the observed flux in black, the estimated continuum emission in red and the corresponding emission lines in blue. The forest of Ly*α *absorption is visible blueward from the Ly*α *emission line.*

scattered away and is called a Lyman absorption because energy is missing at the corresponding wavelength.

If one assumes that the physical state of the absorbing medium is uniquely defined by its temperature and column density, single-absorption line profiles are ideally modeled by a Voigt profile. It is described by the convolution of a Gaussian and a Lorentzian distribution function, known as a Voigt-Hjerting function (Hjerting, 1938).

First, the scattering cross-section for the Lyα transition of an individual atom at rest is modeled by the Lorentzian-profile function of frequency $\nu = 1/\lambda$, which gives the intrinsic width due to the finite lifetime in an excited state,

$$\sigma_{\mathrm{intrinsic,Ly}\alpha}(\nu) = \frac{\pi e^2}{m_e c}\frac{1}{4\pi\epsilon_0} f_{\mathrm{Ly}\alpha}\frac{\Gamma/4\pi^2}{(\nu-\nu_{Ly\alpha})^2+(\Gamma/4\pi)^2} \tag{2.1}$$

where e is the electronic charge, m_e is the mass of the electron, $f_{\mathrm{Ly}\alpha}$ is the oscillator strength that measures the quantum mechanical departure from the classical harmonic oscillator, Γ is a damping term and $\nu_{\mathrm{Ly}\alpha} = 1/\lambda_{\mathrm{Ly}\alpha}$.

In general, the atoms are not at rest. They undergo thermal motions described by a Maxwellian-velocity distribution corresponding to their temperature T, which we model by convolving the intrinsic-line profile (equation (2.1)) with the Gaussian profile produced via the Doppler effect of thermal motions. To account for it we use the Hjerting function,

$$\phi_{\nu,\mathrm{V}} = \frac{1}{\pi^{1/2}}H(r, x(\nu)) = \frac{r}{\pi^{3/2}}\int_{-\infty}^{\infty}\frac{e^{-y^2}}{(x-y)^2+r^2}dy, \tag{2.2}$$

where $r = \frac{\Gamma}{4\pi\Delta\nu_{\mathrm{D}}}$ is the ratio of the damping term to the Doppler width, $x = (\nu-\nu_{\mathrm{Ly}\alpha})/\Delta\nu_{\mathrm{D}}$ is the frequency shift from line center and $\Delta\nu_{\mathrm{D}} = b\nu_{\mathrm{Ly}\alpha}/c$. The Doppler parameter b accounts for thermal motion, it is defined by

$$b = \left(\frac{2k_BT}{m_H}\right)^{1/2}, \tag{2.3}$$

with k_B the Boltzmann constant, m_H the mass of the hydrogen atom and T its temperature. The Doppler width is often much larger than the natural width given by the Lorentzian profile (equation (2.1)), so we can neglect the latter in the following. We then get a profile that is almost exactly Gaussian in the center, but with $1/\nu^2$ wings at the extremities.

Finally, the Lyα resonance-line scattering cross section is given by the following Voigt profile

$$\sigma_{\mathrm{Ly}\alpha}(\nu) = \frac{\pi e^2}{m_e c}\frac{1}{4\pi\epsilon_0} f_{\mathrm{Ly}\alpha}\frac{1}{\Delta\nu_{\mathrm{D}}}\phi_{\nu,\mathrm{V}}. \tag{2.4}$$

For typical densities and temperature in the IGM we can use the Voigt-Hjerting function to zeroth order, which is valid for $r \ll 1$, i.e. when the damping width is negligible compared to the Doppler width. The line profile can then be modeled by the Gaussian core of the Voigt profile,

$$\sigma_{\mathrm{Ly}\alpha}(\nu) = \frac{\pi e^2}{m_e c}\frac{1}{4\pi\epsilon_0} f_{\mathrm{Ly}\alpha}\frac{1}{\Delta\nu_{\mathrm{D}}}\phi_{\nu,\mathrm{G}}, \tag{2.5}$$

with the Gaussian profile

$$\phi_{\nu,\mathrm{G}} = \frac{1}{\pi^{1/2}}e^{-x^2}. \tag{2.6}$$

From these equations arises the classification of the intervening gas presented in Sec. 2.1.2.2. For column densities $N_{\mathrm{HI}} > 10^{20.3}\mathrm{cm}^{-2}$, we cannot ignore the damping width that dominates over the Gaussian core producing large damping wings. Low column-density absorptions can be modeled by the Gaussian profile of equation (2.5), but denser systems such as LLS or DLAs must resolve the full Voigt profile of equation (2.4).

Using only astrophysical arguments, the modeling of Lyα-line profile relies on three parameters ; the position, the column density and the line width that is directly related to velocity dispersion. Line-parameters fitting is used to infer basic physical properties of absorbers, such as the redshift evolution of the column densities, upper limits on the gas temperature or the size and density of Lyα clouds. Furthermore, accurate modeling is fundamental to confront observations with theoretical predictions. I will use the Gaussian-line profile to model Lyα lines in hydrodynamical simulations in Chapter 6.

2.2.2 The Gunn-Peterson effect

With light from a distant quasar serving as a background source, the intrinsic flux get scattered by clouds of hydrogen along the LOS at the observed wavelength λ defined by

$$\lambda = (1 + z_{\mathrm{Ly}\alpha})\lambda_{\mathrm{Ly}\alpha}, \tag{2.7}$$

where $z_{\mathrm{Ly}\alpha}$ is the redshift of the absorber. Because of the cosmological expansion, the rest-frame wavelength of the Lyα absorption gets redshifted as the quasar signal travels in the universe. Hence, the authors in Gunn and Peterson (1965) first predicted (also independent predictions from other groups (Scheuer, 1965; Shklovskii, 1965)) that any neutral clouds of hydrogen along the trajectory of photons from quasars should produce a plethora of absorption lines in their spectra, between the Lyα and Lyβ emissions, which are known today as the Lyα forest. The mechanism is illustrated in Fig. 2.6. Let us consider a quasar at redshift $z_q = 3$ and a cloud of intergalactic hydrogen on the LOS at $z_{\mathrm{Ly}\alpha} = 2.8$. The Ly$\alpha$ emission peak is observed at $\lambda_{\mathrm{e}} = \lambda_{\mathrm{Ly}\alpha}(1 + z_{\mathrm{q}}) = 4860\text{Å}$ but the Lyα absorption caused by the intervening system is observed at $\lambda_{\mathrm{a}} = 4617$ Å, which is on the left of the emission peak.

One can also define other Lyman forests depending on the electronic transition at play, like the Lyβ forest which starts below the Lyβ emission peaks ($\lambda_{\mathrm{Ly}\beta} = 1026\text{Å}$). Ly$\alpha$ absorptions can still occur redwards of the Lyβ emission lines. However, it becomes intermingled with the Lyβ absorption's own forest.

The intergalactic origin of the Lyα forest was demonstrated by the analysis of six quasar spectra by Sargent et al. (1980). The authors showed a strong homogeneity in the properties of the absorbers and through redshift, proving that the signal must come from intervening material, rather than material ejected by the quasars. Also the measured velocity widths corresponded to cloud of gas with temperature $T \sim 10^4$K, which is the characteristic temperature for primordial gas with no cooling via metals.

The observed flux of a quasar, f_{obs}, is related to its intrinsic flux, f_{QSO}, by

$$f_{\mathrm{obs}}(\lambda) = e^{-\tau(\lambda)} f_{\mathrm{QSO}}(\lambda) \tag{2.8}$$

where τ is the optical depth due to the absorption by neutral hydrogen or metals. Using the Beer-Lambert law, the Lyα optical depth is defined by

$$\tau_{\mathrm{Ly}\alpha}(\lambda) = \int_0^{r_{\mathrm{QSO}}} n_{\mathrm{HI}}(r)\sigma_{\mathrm{Ly}\alpha}(\lambda, r)\mathrm{d}r, \tag{2.9}$$

with r the radial distance along the LOS, n_{HI} the column-neutral density and $\sigma_{\mathrm{Ly}\alpha}$ is the Lyα-scattering cross section (see equation (2.5)). The latter depends on the observed wavelength

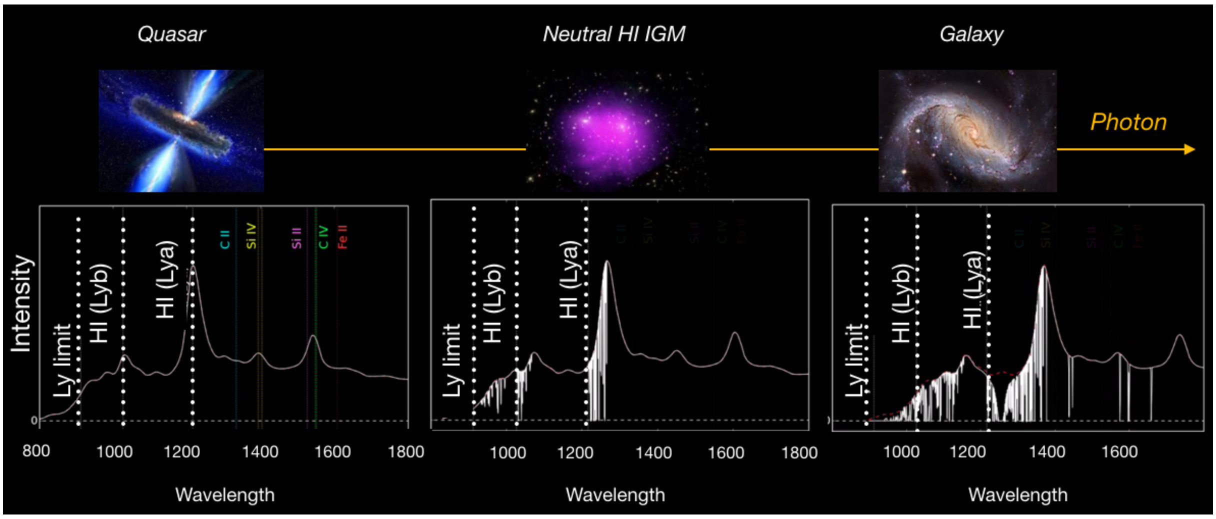

Figure 2.6 – *Illustration of the impact of intervening clouds on a quasar spectrum in an expanding universe. I show the light-flux spectrum along the line of sight : in the quasar vinicity (left), in the IGM (middle) and in the vinicity of a galaxy (right). The spectra are shown in the rest-frame of the light flux. As it travels in space, the spectra gets redshifted. Hence, each neutral hydrogen absorptions (here I show Lyα and Lyβ absorptions and emissions) appear blueward from the associated emission line. The Lyα absorption associated to the one in the vinicity of the galaxy produces a saturated absorption clearly distinguishable from the low-density absorptions in the IGM. Metal absorption lines (redward from the Lyα forest) only appear traveling near a galaxy, because the IGM is mainly composed of primordial gas. Adapated from* Andrew Pontzen's video .

λ and on the radial distance r because of the cosmological expansion redshifting the absorbed wavelength. The path length in the sightline is $\mathrm{d}r = a\mathrm{d}x = \mathrm{d}v/H$ where r is the proper distance, x is the comoving distance and v is the Hubble flow velocity. However, when we measure a redshift distance we really measure a velocity. Peculiar velocities can alter this recessional velocity that is expected from the Hubble flow and modifies the observed redshift via the relativistic Doppler effect. Therefore, peculiar velocities of the intervening systems must be taken into account because they shift the position of the absorption lines. In velocity coordinate the optical depth is

$$\tau_{\mathrm{Ly}\alpha}(v) = \frac{\pi e^2}{m_e} \frac{\lambda_{\mathrm{Ly}\alpha}}{4\pi\epsilon_0} f_{\mathrm{Ly}\alpha} \int n_{\mathrm{HI}}(v') \frac{1}{\pi^{1/2} b(v') H} e^{-\left(\frac{v - v_0(v')}{b(v')}\right)^2} \mathrm{d}v', \tag{2.10}$$

where $v_0(v') = v' + v_{||}(v')$ is the line-of-sight velocity of gas particle, with $v_{||}(v')$ the peculiar velocity of the particle along the LOS. The process of going from baryon density to flux taking into account all the above effects is shown in Fig 2.7. The middle panel illustrates how the velocity component along the LOS shift the position of the line and shows that redshift-space distorsions also modify the blending of the lines.

Based on the assumption that the gas is in photoionization equilibrum (but not necessarly in thermal equilibrium) with the cosmic UV background field, i.e. the UV background is balanced by the electron recombination,

$$\gamma + \mathrm{H} \leftrightarrows \mathrm{e}^- + \mathrm{H}^+, \tag{2.11}$$

the neutral hydrogen density is proportional to the baryon density with $n_{\mathrm{HI}} \propto T^{-0.7} n_{\mathrm{b}}^2$, where T is the temperature of the gas. For intergalactic gas, the interplay between photoionization hea-

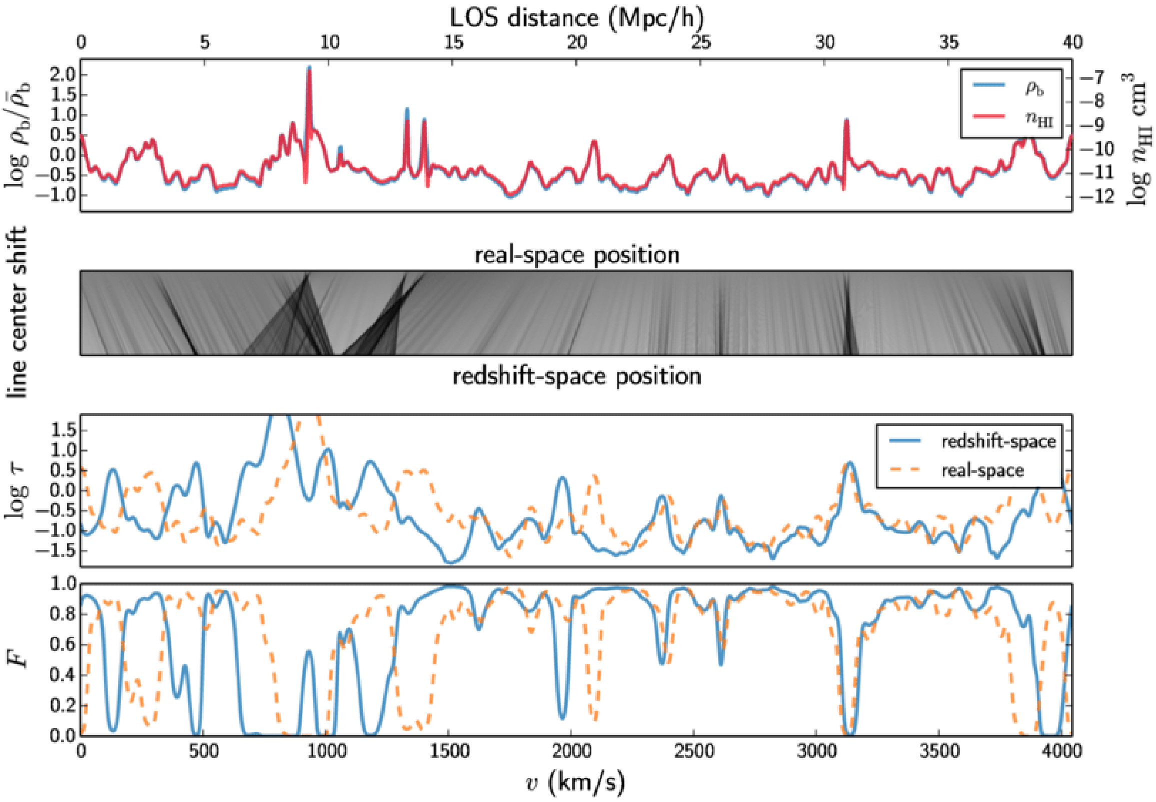

FIGURE 2.7 – *The process of going from baryon density to flux spectra on a simulated line of sight (LOS). The top horizontal axis gives the comoving spatial position along the LOS in* Mpc · h^{-1} *and the lower horizontal axis gives the coordinate value in redshift space in* km · s^{-1}*. The top panel gives the baryon density, the middle panel illustrates how the velocity component along the LOS shift the position of the line and shows that redshift-space distorsions also modify the blending of the lines. The lower two panels show the optical depth and flux along the LOS. From Lukić et al. (2015).*

ting, recombination cooling and adiabatic cooling leads to a tight temperature-density relation

$$T = T_0(1 + \delta_b)^{\gamma-1}, \tag{2.12}$$

where T_0 and γ are constants that depend on the thermal history, and δ_b is the baryon-density fluctuations. With this physical reasoning, and by neglecting thermal broadening and peculiar velocities, the optical depth can be converted in a formula described as the *fluctuating Gunn-Peterson approximation* (FGPA) (Weinberg et al., 1997; Croft et al., 1998),

$$\tau_\mathrm{FGPA} \propto 172 \left(1 + \frac{v_{||}}{H(z)}\right)^{-1} \left(\frac{1+z}{4}\right)^{6} \left(\frac{H_0}{H(z)})\right) \left(\frac{\Omega_\mathrm{b} h^2}{0.0125}\right)^{2} \left(\frac{T_0}{10^4\mathrm{K}}\right)^{-0.7} (1+\delta_\mathrm{b})^{\beta}, \tag{2.13}$$

where $\beta = 2 - 0.7\gamma$. This approximation breaks down above $\delta_\mathrm{b} \geq 10$ and at low redshifts. The simplicity of the physics governing the IGM makes the FGPA a very powerful tool to estimate the general abundance of neutral hydrogen in the universe. Because the quasar flux depends exponentially on the Lyα optical depth, which in turn depends almost quadratically on the hydrogen density, small density fluctuations produce enhanced fluctuations in the transmitted flux. Moreover, the Lyα cross section is so large that even if only a small fraction of the mass in the universe was in the form of neutral hydrogen, the redshifted Lyα absorption would produce a discontinuity in the spectrum just after the Lyα emission.

However, Gunn and Peterson (1965) only reported a small decrement in the observed flux of quasar. It either meant that galaxy formation converted almost all hydrogen into stars and very little primordial hydrogen was left in the intergalactic medium, or that most of the hydrogen was ionized. Quasar surveys later confirmed the second hypothesis by showing that the total ionizing radiations from quasars was strong enough to keep most of the baryonic matter in the universe highly ionized.

2.2.3 A signature of cosmological structure formation

The decomposition of the Lyα forest into Voigt profiles was justified by the assumption that such absorption features arise from distinct clouds in hydrostatic, thermal and ionization equilibrium, pressure-confined by a hot IGM (Sargent et al., 1980). However this model suffered from many drawbacks in contradiction with observational evidences. Observations of the CMB by the COBE satellite provided evidence against a hot IGM necessary to provide the pressure-confinement of the Lyα clouds. Moreover, such clouds are expected to have typical sizes of 17 kpc, in strong disagreement with sizes inferred from coincident absorbers along neighboring LOS that give cloud sizes exceeding 70 kpc (Smette et al., 1992; Bechtold et al., 1994; Dinshaw et al., 1997).

The success of the cold dark matter dominated scenario to describe the large-scale distribution of galaxies (Davis et al., 1985) led to attempt to describe the Lyα forest within this structure-formation model. The cold dark matter model predicts an abundance of low-mass minihalos, too small to form stars, that would gravitationnaly confine the neutral clouds in the IGM (Rees, 1986; Ikeuchi, 1986). The intergalactic gas falls into these minihalos and the thermal pressure prevents them from further collapsing leading to relatively stable gaseous configurations (Bond et al., 1988). We can only detect these structures in absorption since there are no stars to produce light. The first hydrodynamic cosmological simulations with sufficient resolution and dynamic range to model the intergalactic gas showed that the Lyα forest is a natural consequence of the growth of structure in the cold dark matter hierarchical-gravitational-collapse scenario by successfully reproducing its basic properties (Cen et al., 1994; Zhang et al., 1995; Katz et al., 1996; Theuns et al., 1998). In this underlying cosmological model, hydrodynamical simulations showed that the IGM is arranged in filaments and sheets, forming the now well-known *cosmic web*. Also, they showed that IGM is the original reservoir of matter for the formation of galaxies. As such, the reproductibility of the Lyα forest constitutes one of the strongest observational evidence for hierarchical structure formation.

2.3 The Lyα forest in cosmology

Numerical simulations have then proved that the Lyα forest arises from density and velocity fluctuations of a smooth IGM. At high redshifts ($z \geq 2$), most of the intergalactic gas has experienced only mild gravitational collapse, so that absorption lines are near the linear regime with typical density fluctuations of $\delta \leq 10$. Its dynamics at first glance appears to be quite simple with gravitational instabilities driven by dark matter fluctuations on large scales and baryonic-pressure smoothing on small scales (Reisenegger and Miralda-Escude, 1995; Bi and Davidsen, 1997; Hui et al., 1997). Therefore, dark matter and baryon components trace each other on large scales, even better than boolean tracers like galaxies or quasars. Hence, the Lyα forest is sensitive to cosmological parameters and can be exploited to constrain cosmological models by comparing observations and theoretical predictions from hydrodynamical simulations. For instance, comparing the absorption-line parameters, such as the Doppler-width distribution (the b-distribution) (Hu et al., 1995; Lu and Haung, 1996; Theuns et al., 1998; Schaye et al., 1999; Bolton et al., 2014; Garzilli et al., 2015; Rorai et al., 2018; Hiss et al., 2019), efficiently constrain thermal history and reionization models. However, estimating line parameters in numerical simulations is computationally demanding. More straightforward comparison can be made using direct flux statistics.

The Lyα flux autocorrelation function, $\xi_{\mathrm{Ly}\alpha}(\mathbf{r})$, introduced by Zuo and Bond (1994) is defined by analogy with the autocorrelation function of galaxies,

$$\xi_{\mathrm{Ly}\alpha}(\mathbf{r}) = \langle \delta_{\mathrm{Ly}\alpha}(\mathbf{x}')\delta_{\mathrm{Ly}\alpha}(\mathbf{x})\rangle_{\mathbf{r}} \,, \tag{2.14}$$

where $\delta_{\mathrm{Ly}\alpha}$ is defined as the transmitted Lyα flux fraction. The latter describes the fluctuations of the Lyα transmitted flux around its mean value, such that

$$\delta_{\mathrm{Ly}\alpha}(\mathbf{x}) = \frac{f(\mathbf{x})}{\langle f \rangle}, \tag{2.15}$$

with $f(\mathbf{x})$ the pixel flux at coordinate $\mathbf{x}$. I stress that $\delta_{\mathrm{Ly}\alpha}$ is the transmitted flux fraction of Lyα absorptions only. Indeed, in observed spectra we also detect absorptions caused by metals or by other electronic transitions from neutral hydrogen such as the Ly-β absorption. Thus, we can equivalently define the total transmitted flux fraction δ_{F} that accounts for all absorptions. Several estimates using pixel information from different quasar spectra have been measured (Liske et al., 2000; McDonald et al., 2000; Croft et al., 2002; Becker et al., 2004; Kirkman et al., 2007; Slosar et al., 2011; Busca et al., 2013; Slosar et al., 2013; Delubac et al., 2015; Bautista et al., 2017; de Sainte Agathe et al., 2019). Equivalently, flux cross-correlations with other matter tracers such as galaxies, quasars or DLAs can be measured (Ryan-Weber, 2006; Font-Ribera et al., 2014; du Mas des Bourboux et al., 2017; Pérez-Ràfols et al., 2017). It ideally probes scales of tens to hundreds of comoving Mpc. Hence, it is powerful to constrain the expansion of the universe and the dark energy content of the universe by measuring the position of the BAO peak at high redshifts, which cannot be done with other tracers. Theoretically, it can also be used to detect the redshift distorsions predicted in linear theory of large-scale structures by gravitational evolution (Kaiser, 1987).

The Fourier counterpart of the autocorrelation function is called the 3D Lyα power spectrum, and is defined by

$$P_{\mathrm{Ly}\alpha,\mathrm{3D}}(\mathbf{k}) = \tilde{\xi_{\mathrm{Ly}\alpha}}(\mathbf{k}), \tag{2.16}$$

where $\mathbf{k} = 2\pi/\mathbf{r}$ is the wavenumber. By analogy with the 3D matter power spectrum, the 3D Lyα power spectrum is also the measure of the variance in the amplitude of the Fourier transform coefficients of the transmitted Lyα flux $\delta_{\mathrm{Ly}\alpha}$,

$$P_{\mathrm{Ly}\alpha,\mathrm{3D}}(\mathbf{k}) = \left| \left\langle \tilde{\delta_{\mathrm{Ly}\alpha}}(\mathbf{k}) \right\rangle \right|^2 . \tag{2.17}$$

Power spectrum and autocorrelation function are simple Fourier pairs, containing possibly the same information. However they are estimated using different algorithms, sensitive to different systematics and are not compared directly. Methods and algorithms have been developped to estimate the $P_{\mathrm{Ly}\alpha,\mathrm{3D}}(\mathbf{k})$ (Font-Ribera et al., 2018), but have never been used with real data yet because a high-density of close quasar spectra is required.

Finaly, by using pixel information on same LOS only, we can define the 1D Lyα power spectrum by

$$P_{\mathrm{Ly}\alpha}(k_{||}) = \frac{1}{(2\pi)^2} \int P_{\mathrm{Ly}\alpha,\mathrm{3D}}(\mathbf{k}) d\mathbf{k}_{\perp}, \tag{2.18}$$

where the wave number $\mathbf{k} = (k_{||}, \mathbf{k}_{\perp})$ is decomposed in the radial and transverse directions. Equivalently, if we measure the Lyα transmitted flux field $\delta_{\mathrm{Ly}\alpha}(x_{||})$ in the LOS direction only, i.e. radial direction, we have,

$$P_{\mathrm{Ly}\alpha}(k_{||}) = \left| \left\langle \tilde{\delta_{\mathrm{Ly}\alpha}(k_{||})} \right\rangle \right|^2 . \tag{2.19}$$

These last statistics have received a lot of attention thanks to its apparent simplicity, and have been measured on different data sets (Croft et al., 1998, 1999; Kim et al., 2004; Viel et al., 2004; McDonald et al., 2006; Viel et al., 2008; Palanque-Delabrouille et al., 2013; Viel et al., 2013; Iršič et al., 2016; Yèche et al., 2017; Walther et al., 2018). It is an invaluable cosmological probe because it traces density fluctuations down to the Mpc scale in a unique redshift range with $2 \leq z \leq 6$. As such, it opens a window on a unique range of scales unvailable before.

Two broad approaches have been pursued to constrain cosmology with $P_{\mathrm{Ly}\alpha}$. In the first one, measurements are inverted to recover the underlying dark matter power spectrum (Croft et al., 1998; Hui et al., 1999; Nusser and Haehnelt, 1999). In particular, it has been used as the smallest-scale probe for the estimate of the matter power spectrum spanning cosmic times and cosmic scales in Tegmark and Zaldarriaga (2002), shown in Fig. 1.8. In the second approach, we use directly the flux power spectrum to infer cosmological parameters such as the scalar spectra index n_s, the amplitude of mass fluctuations σ_8 or the density of matter Ω_m. More specifically, because the $P_{\mathrm{Ly}\alpha}$ is sensitive to clustering on the smallest scales, it is sensitive to the smoothing of relativistic particles. Fig. 1.9 shows the impact of massive neutrinos on the matter power spectrum for different masses and different redshift. The yellow band indicates the range of scales probed by the Lyα forest, which ideally probes the maximal suppression region.

I will use these two approaches in Chapter 7 by exploiting the 1D power spectrum measurements described in Chapter 4, in combination with other cosmological probes, to infer the matter power spectrum at $z = 0$, to constrain the sum of neutrino masses and determine the plausibility of a warm dark matter model.

I want to stress that the Lyα forest is not just an additonal tool to confirm already existing results. Indeed, the 1D power spectrum of the Lyα forest probe scales significantly smaller than other probes, as it can go below the Mpc scale with high-resolution Lyα surveys, such as the one using the High Resolution Echelle Spectrometer (HIRES) or the Magellan Inamori Kyocera Echelle (MIKE) spectrometer. Also, it opens a window on a range of redshifts where the evolution of fluctuations are not erased by non linearities yet. As such, the Lyα forest is an invaluable tool at measuring the smoothing of fluctuations induced by the free streaming of relativistic particles. In particular, it has been used to put the strongest constraints to date on the mass of hypothetical warm dark matter or fuzzy dark matter particles (Baur et al., 2016; Armengaud et al., 2017; Yèche et al., 2017; Iršič et al., 2017; Baur et al., 2017).

In addition, because it probes scales and redshifts perpendicular to other probes, it allows to constrain cosmology in a holistic way. First, when combining with other probes, the Lyα forest can **break degeneracies and significantly thighten constraints** because of the information embedded in these smallest scales and because it is sensitive to different systematics. For instance, neutrinos leave a signature in the CMB power spectrum through the integrated

Sachs-Wolf effect and through lensing (Lesgourgues et al., 2013; Planck Collaboration et al., 2018a) at the level of ~ 0.5 eV. Lyα data alone has sensitivity to the sum of neutrino masses at the level of ~ 0.6 eV since the Lyα forest probes scales where the ratio of the power spectra for massive to massless neutrinos is quite flat. A tight constrain at the level of ~ 0.1 eV can be obtained by combining CMB and Lyα since the combination probes the power spectrum both in the suppressed and unaffected regions. Thus, Lyα measures the power spectrum level, defined by σ_8 and Ω_m, CMB provides the correlations between these parameters and $\sum m_\nu$, and the joint use of these two probes significantly improves the constraint on $\sum m_\nu$ compared to what either probe alone.

Then, Lyα analysis can **reveal tensions** when confronted with different and independent probes, which potentially brings to light the need of new physics beyond the standard model. For instance, Lyα constraints present tensions on n_s, or equivalently on Ω_m, when compared to CMB data (Palanque-Delabrouille et al., 2015,?; Yèche et al., 2017). It is of particular interests since the well-known σ_8 tension between weak-lensing surveys and CMB can be equally interpreted as an Ω_m tension that goes in the same direction that the Lyα/CMB tension. The fact that Lyα and weak-lensing data are two late time probes of a similar range of scales that agree with each other while they are both in tension with the CMB early time probe, potentially indicates a failure in the standard cosmological model. This will also be investigated in Chapter 7.

3

The extended Baryon Oscillation Sky Survey

Contents

THE Sloan Digital Sky Survey (SDSS) is one of the largest, most ambitious and influential surveys in the history of astronomy, with the goal of pushing further our knowledge and understanding of the large-scale structure evolution, the formation of stars and galaxies including the Mikly-Way and the source of dark energy. It consists in a major multi-spectral imaging and spectroscopic redshift survey using a dedicated 2.5-meter telescope at the Apache Point Observatory (APO) in New-Mexico. By thoroughly mapping over a third of the night sky, the SDSS represents one of the major quests of contemporary physics and has driven discoveries on many fundamental questions about the origins of the universe. Studies show that SDSS is among the most highly-used and cited data sets in astronomy. SDSS data has been used in over 7,800 papers and 390,000 citations, 80 percent of which have been published by scientists outside the SDSS collaboration. The SDSS project was imagined in the eighties by Jim Gunn, Rich Kron and Donald York. The Sloan foundation played an important role in supporting the project, which has now over 54 university partners who contribute most of the funding for SDSS. In addition to significantly accelerating discoveries, the SDSS transformed how astronomical research gets done thanks to its open data principles. It is a milestone for the development of international collaboration in astronomy. This work uses a 6-year data sample of high redshift quasars from the BOSS and eBOSS programs, two surveys of the third and fourth generation of SDSS. Sec. 3.1 provides a brief overview of the different generations of the SDSS and their main surveys. Then, Sec. 3.2 presents the main technical characteristics of the SDSS telescope, the camera and spectrographs that were designed for BOSS and also used for eBOSS. Finally, Sec. 3.3 presents the main data reduction steps, including characteristics of the pipeline software, visual inspection strategy and the automated detection of strong absorbers.

3.1 The Sloan Digital Sky Survey

The survey is divided in four phases, each of them represents about five years of observations, answering different key questions in astronomy. The fifth generation will begin observations in Summer 2020. I will here briefly describe the major scientific goals and results of the previous generations, in particular the third and fourth ones, from which I used data to perform the Lyα forest data analysis.

3.1.1 SDSS-I and II

SDSS first light was observed in 1998, then observations started in 2000. The first and second generations of the SDSS survey (SDSS-I and SDSS-II) were performed during 2000-2005 and 2005-2008 respectively. During its first phase operation, the SDSS produced a multi-band photometric survey of more than $8,000\,\mathrm{deg}^2$ of the sky along with a spectroscopic survey of galaxies and quasars of $5,700\,\mathrm{deg}^2$ of the sky included in the imaging survey.

During its second phase operation, the SDSS produced three major surveys :

- The **Sloan Legacy Survey** completed the footprint observed by SDSS-I and produced an imaging and spectroscopic survey covering $8,400\,\mathrm{deg}^2$ of the Northern Galactic Cap including photometric data of 2 million objects and spectroscopic data of 800,000 galaxies and 120,000 quasars. The information on the position and distance of the objects has allowed the large-scale structure of the Universe, with its voids and filaments, to be investigated for the first time.

- The **Sloan Extension for Galactic Understanding and Exploration** (SEGUE) created a detailed three-dimensional map of the Milky Way that provided clues for the formation and evolution of our galaxy through a kinematic and stellar population study of the high-latitude thick disk and halo of the Milky Way. It provided photometric data of $3,500\,\mathrm{deg}^2$ of the sky, along with the spectra of 240,000 stars.
- The **Sloan Supernova Survey** produced a repeat imaging of the $300\,\mathrm{deg}^2$ southern equatorial stripe to detect variable objects and supernovae (SN). It discovered almost 500 spectroscopically confirmed Type 1a SN that are particularly useful to determine the accelerating cosmic expansion of the last billion years.

All the imaging, spectroscopic data and derived parameters are publicly available with the seventh data release (DR7) and can be found on the SDSS website [1]. The study of these data had a tremendous impact on our understanding of structure formation in the universe either at the cosmological scale or the astrophysical scale. In particular, the SDSS achieved the first clear detection of the BAO signal in the two-point correlation function of a sample of 46,000 luminous red galaxies (LRG) at en effective redshift of 0.35 with a precision of 5% (Eisenstein et al., 2005).

3.1.2 SDSS-III

The third generation of SDSS [2] started in Automn 2008, directly following SDSS-II. It used the same telescope but with significant improvements on the doubled-armed spectrographs (later described in Sec. 3.2). The final data were publicly release in DR12 [3]. With the aim to improve our understanding of the dark energy, to map the Milky-Way and to search for extra-solar giant planets, SDSS-III is built on four different surveys :

- The **Baryon Oscillation Spectroscopic Survey** (BOSS) was a six-year spectroscopic program designed to refine the measurement of the standard ruler with enough sensitivity to constrain both the expansion rate and the angular diameter distance to the percent level precision. To achieve this, BOSS carried out two programs on $10,000\,\mathrm{deg}^2$ of the sky. The first one consists in the observation of 1.5 million LRG with $0.15 \leq z \leq 0.7$. The second one is dedicated to the observation of 160,000 high redshift quasars to constrain the BAO scale at $z \sim 2.5$ with the Lyα forest.
- The **Sloan Extension for Galactic Understanding and Exploration 2** (SEGUE) doubled the sample size of SEGUE-1 by producing the spectra of 120,000 stars, focusing on the in situ stellar halo of the Milky Way. Combined, SEGUE-1 and SEGUE-2 revealed the complex kinematic and chemical substructure of the Galactic halo and disks, providing essential clues to the assembly and enrichment history of the Milky-Way.
- The **Apache Point Observatory Galactic Evolution Experiment** (APOGEE) observed with high precision the peculiar velocity and the chemical composition of 100,000 very bright red giant stars in different regions of the Milky Way using high-resolution infrared spectroscopy. It revealed the abundance of about 15 chemical elements giving clues on the dynamical structure and chemical history of the Milky Way.
- The **Multi-object APO Radial Velocity Exoplanet Large-area Survey** (MARVELS) was a spectroscopic survey designed to observe 11,000 bright stars of the Milky Way. Each star was to be observed multiple times to constrain theoretical models of

---

[1] https ://classic.sdss.org/dr7

[2] http ://sdss3.org/

[3] http ://sdss3.org/dr12

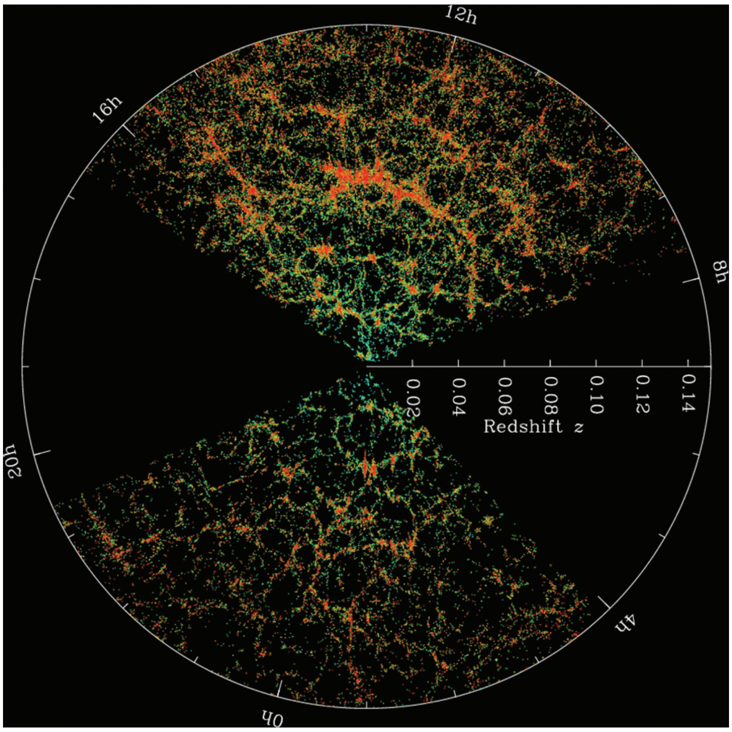

Figure 3.1 – *The SDSS's map of the universe. Credit : M. Blanton and SDSS*

formation and evolution of giant planet systems. However, the required resolution of the spectra was never reached and the project was stopped in 2012.

By the end of SDSS-III, the wealth of provided data allowed the study of the cosmic web, as shown in Fig. 3.1, where clusters and walls of galaxies, which are the largest structures in the entire universe, are visible.

3.1.3 SDSS-IV

The fourth generation of SDSS [4] lasted from Automn 2014 to February 2019, all data were made public in DR16 [5]. With the aim to push even further the precision of cosmological measurements on dark energy, to expand its infrared spectroscopic survey of the Galaxy in the northern and southern hemispheres and to make spatially resolved maps of individual galaxies, SDSS-IV is built on three surveys :

- The **extendend Baryon Oscillation Spectroscopic Survey** (eBOSS), in the continuity of BOSS, measures the expansion history of the Universe throughout eighty percent of cosmic history, back to when the Universe was less than three billion years old. It aims at probing the transition from deceleration of the cosmic expansion, at testing general relativity and modified theories of gravity and at detecting the effects of cosmic neutrinos. It includes four classes of targets : a sample of LRG at higher redshifts than the BOSS sample, a sample of emission line galaxies (ELG), i.e. blue and star forming galaxies, with $0.6 \leq z \leq 1.2$, a sample of low redshift quasars with $0.8 \leq z \leq 2.2$ to study quasar clustering and a sample of high redshift quasars with $z > 2.2$ to study Lyα forest cluste-

---

[4] https ://www.sdss.org/

[5] https ://www.sdss.org/dr16/

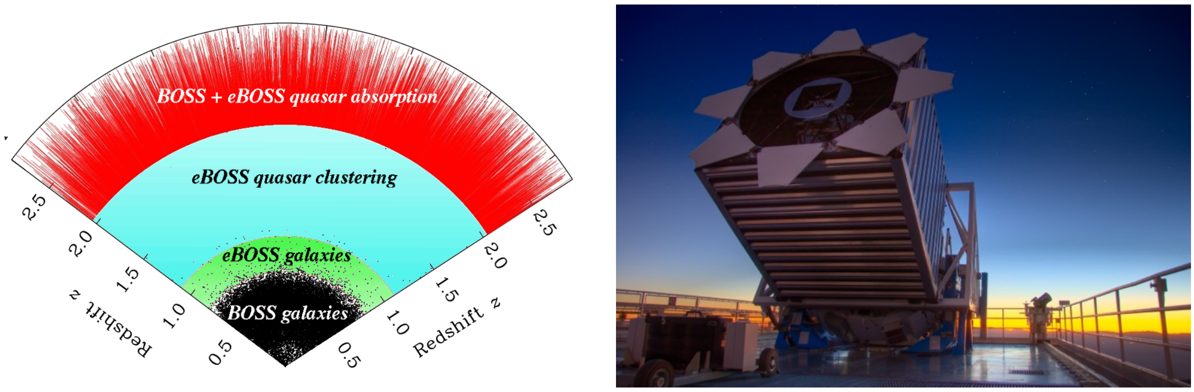

FIGURE 3.2 – ***Left :*** *BOSS and eBOSS redshift coverage for the different targets. Credits : SDSS eBOSS* ***Right :*** *The 2.5m SDSS telescope at the Apache Point Observatory in New-Mexico (USA). Credits : SDSS*

ring. The latter is used in this thesis and the next Chapter is dedicated to the analysis of this Lyα sample. The distribution in redshifts of the different targets is shown in Fig. 3.2 (left panel). Details about the science forecasts can be found in Dawson et al. (2016); Zhao et al. (2016). eBOSS also includes two subprograms to follow up on other types of objects : the Time-Domain Spectroscopic Survey (TDSS) for variable objects, and the SPectroscopic IDentification of ERosita Sources (SPIDERS) for X-ray sources.

- **APOGEE-2** continues to probe the history of star formation and chemical enrichment of the Milky Way along with the dynamics of its disk, bulge, and halo. In addition to the SDSS telescope, APOGEE-2 also uses the 2.5-meter du Pont Telescope at Las Campanas Observatory in Chile to extend its observations to the Southern hemisphere component to look for previously unreachable parts of the disk. Along with red giant star spectra, APOGEE-2 acquires spectra of young stars and star-forming regions, variable stars, stars in clusters and satellite galaxies, and stars with asteroseismic measurements.
- The **Mapping Nearby Galaxies at APO** (MaNGA) survey aims at understanding the history of present day galaxies from imprinted clues of their birth and assembly, by providing two-dimensional maps of stellar velocity and velocity dispersion, mean stellar age and star formation history, stellar metallicity, element abundance ratio, stellar mass surface density, ionized gas velocity, ionized gas metallicity, star formation rate and dust extinction.

3.2 The Instrument

A full description of the technical characteristics of the instrument can be found in York et al. (2000) and in Gunn et al. (2006). I give here a brief summary of observational parameters and data products of the SDSS, which are necessary for the Lyα forest data analysis described in Chapter 4.

The SDSS was imagined as a multi-band photometric survey covering a large fraction of the observable sky associated with an ambitious spectroscopic survey. Such requirements imply several technical innovations for each element.

- A **telescope** with a wide field of view, a very low distorsion focal plane, a very precise pointing and a capacity to switch easily between photometry and spectroscopy

- A **camera** with a wide field of view, able to observe simultaneously in different photometric bands and a good precision for astrometric calibration
- Two **spectrographs** able to take hundreds of spectra simultaneously with a wide wavelength range
- An **acquisition system** with unprecedented data storage capabilities and real time control quality
- A fast **data pipeline** to efficiently calibrate the data and identify objects both in photometry and spectroscopy

The two key technologies that enabled the SDSS, optical fibers and the digital imaging detectors known as CCDs, were the discoveries awarded the 2009 Nobel Prize in Physics.

3.2.1 The Telescope

Requirements for the telescope are achieved with a 2.5m telescope with a 3° field of view, built at the APO in New-Mexico at 2,800m above the sea level, as seen in Fig. 3.2 (right panel). The SDSS telescope is a Ritchy-Chrétien with a primary and secondary hyperbolic mirrors, forming an optical system with a focal to diameter ratio of $f/0.5$. In addition, the telescope includes two optical correctors : a Gascoigne corrector to reduce the optical system astigmatism that is the major drawback of Ritchy-Chrétien telescopes, and a "final" corrector, which is in reality a pair of correctors that are switched depending on the telescope configuration (photometric or spectroscopic).

3.2.2 The Camera

The photometric survey of BOSS and eBOSS used the same camera that was used for SDSS and SDSS-II because it was already optimized for the operating mode and the 3° field of view of the telescope. A full technical description can be found in Gunn et al. (1998). I will briefly summarize here the main elements. The first and most important element is an optical corrector that aims at correcting the distorsions at the focal plane induced by the telescope optical system. It plays a major role in the preservation of the mechanical properties of the camera and has a big impact on the image quality and astrometry. The second element is a matrix of 6 columns of 5 CCDs each, one for each photometric band (u', g', r', i' and z'). The size of one CCD is 2048 × 2048 pixels. The rest of the focal plane is used for 24 smaller CCDs (2048 × 400 pixels) used for focus and astrometry. The CCDs can observe from the UV atmospheric cut off, around 3,000 Å, to the limit of silicon-based detectors close to 11,000 Å.

3.2.3 The Spectrographs

Two identical BOSS spectrographs were rebuilt from the original SDSS spectrographs to take into account upgrades like the increase in the number of fibers or the optical efficiency. They remain the active spectrographs in SDSS and they were used to collect eBOSS and MaNGA data and they will also be the active spectrographs in SDSS-V. I will only recall here the main characteristics (see Smee et al. (2013) for a full technical description).

3.2.3.1 Cartridges and optical fibers

To get the spectra, 1,000 holes are drilled on an aluminium plate , which is 3.2mm thick with 81.3 cm diameter and wheigts 4.3kg. Each hole corresponds to an astronomical object (quasar,

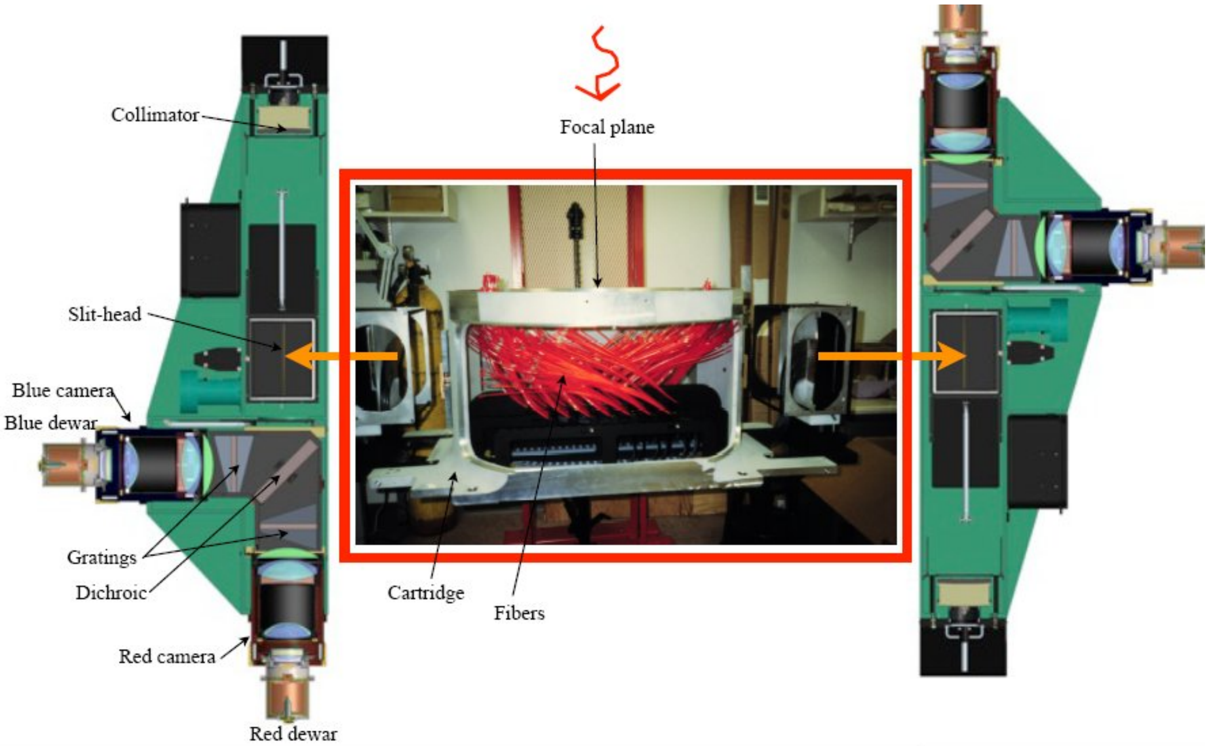

FIGURE 3.3 – *An illustration of the BOSS spectrograph with the aluminium plate at the focal plane, connected to the spectrographs by optical fibers held by the cartridge. The light at the end of the fibers is collected by two collimators, each for 500 fibers, via two slitheads. Each slit goes into a spectrograph when the cartridge is mounted on the telescope. Credits : SDSS*

galaxy, standard star or random blank area for sky and background substraction). The plate must be drilled in advance. Each of them is osberved for typically one hour, hence for one night up to nine plates can be observed. Optical fibers are then plugged into these holes to redirect the light to spectrograph grism. The plates and spectrographs are held by an aluminium structure, the whole forming a cartridge that can be seen in Fig. 3.3.

In order to increase the statistics of the survey compared to the two first generations of SDSS, the number of fibers per plate went from 640 to 1,000. Doing so without changing the telescope optics required to decrease the size of the fibers from 180μm ($3''$ on the sky) to 120μm ($2''$ on the sky).

3.2.3.2 Blue and red cameras

Each spectrograph has two cameras. The blue camera covers $3,600$ Å $\leq \lambda \leq 6,350$ Å, when the initial lower cutoff was $3,900$ Å. This extension allows to increase the detection of the Lyα forest down to $z \sim 2$. The red camera covers $5,650$ Å $\leq \lambda \leq 10,000$ Å, when the initial upper cutoff was $9,000$ Å. This increase was motivated by the study of galaxy spectra discontinuity and is limited by the use of silicon detectors.

The resolution power is defined as $R = \lambda/\Delta\lambda$, where $\Delta\lambda$ is the full width at half maximum of the point spread function of the CCD image by the spectrograph. It is measured before each plate observation using arc lamps dedicated to the calibration. For the blue camera, the resolution power varies between 1,560 and 2,270. For the red camera it varies between 1,850 and 2,650. R

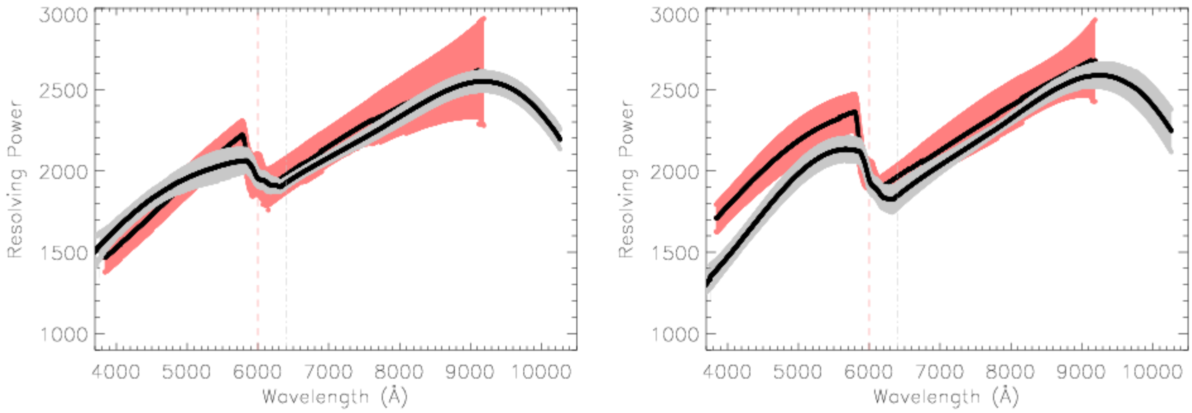

FIGURE 3.4 – *The resolution power for the BOSS spectrographs in grey (spectrograph 1 on the left and spectrograph 2 on the right) and for the original SDSS spectrographs in red. Smee et al. (2013)*

averaged on a hundred plates, as a function of λ is shown in Fig. 3.4. The step around 6,200Å is due to the change of CCDs (blue channel to red channel). R also depends on the position in the focal plane (i.e. the fiber number). For all CCDs the resolution is lower on the sides, at the edge of the spectrographs, by a factor of two compared to the center.

3.3 Data Reduction

The BOSS data reduction pipeline, spec1d, was designed to extract, calibrate, coadd, classify and estimate the redshift of all astronomical objects of each plate using all available individual exposures. The software is fully described in Bolton et al. (2012), I will give here the main steps of the data reduction. However, it suffers from some caveats because of DLAs or BALs in spectra hindering the classification and/or redshift estimation. To overcome this issue visual inspection has been conducted on all quasar spectra up to DR12. In contrast, spectra acquired during the eBOSS survey were not systematically visually inspected.

3.3.1 Pipeline

The extraction step from the raw CCD images transforms the two dimensional images of each individual exposures in one dimensional spectra, which is flux-calibrated using the arc lamps data of the plate. The per pixel variance is estimated using the readout noise and the number of photon recorded for each pixel. The inverse variance is then multiplied by a factor that takes the different known flaws of the CCDs. Pixels that were hit by cosmic ray are identified and masked. Finally, the flux is calibrated using the spectra of standard stars that were observed on the plate for this purpose.

The second step consists in the construction of the coadd spectra combining the different individual exposures. The spectra are then rescaled to be linear in $\log(\lambda)$ with $\Delta(\log(\lambda)) = 10^{-4}$. The variance is obtained using the variance of each exposure, the covariance being neglected. If an object has been observed on multiple nights, the spectrum with the best signal-to-noise is kept as the primary spectrum.

The last step is the classification of the spectra and the estimation of the redshift of the object. The pipeline relies on templates that span both the full space of physical object types within the survey and the full wavelength coverage of the spectrographs. Each template is determined using a training sample of objects with known redshift and whose spectra are decomposed into a common basis of principal components (Principal Component Analysis). This procedure is fully described in Aihara et al. (2011).

3.3.2 Visual Inspection

As shown in Sec. 2.1.2, quasars have a broad variety of spectra characteristics that make their automated classification difficult. To overcome this problem, members from the BOSS collaboration have visually inspected all quasar spectra from the BOSS survey. The visual inspection, fully described in Pâris et al. (2012), allows a reliable identification as well as a precise determination of the redshift. In addition, it allows the detection of particular features such as DLA or BAL.

The visual inspection allows one to compute an efficiency of the automatic classification of the spectra. Star sample appears to be very pure with less than 0.1% of the objects classified by stars being quasars in reality. Quasars are more prone to classification issues, with 0.5% of objects classified as high-redshift quasars ($z > 2.15$) being in reality either stars or lower-redshift quasars. This contamination reaches 5% for low-redshift quasars. The visual inspection also allows testing the efficiency of the redshift estimation of quasars. It showed that only 0.3% of BOSS quasars had an error greater than 0.1 ($\Delta \mathrm{z} > 0.1$). Such errors mainly occured for quasars with a redshift $z < 2.0$ with a non-visible Lyα emission line.

3.3.3 Automated detection of strong absorbers

Since the release of the DR14 quasar catalogue (Pâris et al., 2018), visual inspection is no longer performed on all quasar targets because of the significant increase in statistics. Fully automated and statistically consistent methods have been developed to detect strong absorptions in quasar spectra, which are critical for Lyα forest studies. I will briefly describe the procedures for automated DLA and BAL detection that have been used for Lyα analyses with BOSS/eBOSS spectra.

3.3.3.1 Detection of BALs

BALs are searched for in quasar spectra with $1.57 \leq z \leq 6$ with an algorithm that looks for absorption troughs that would represent blueshifted C IV or Si IV features using a χ^2 fit of an unabsorbed quasar model. The fit is performed in the $1,260\text{Å} \leq \lambda_{\mathrm{RF}} \leq 2,400\text{Å}$ rest-frame window when possible. The algorithm then attempts to measure the commonly used "balnicity index" (BI) proposed by Weymann et al. (1991). BI is computed over the region from 25,000 km/s to 3,000 km/s blueshift relative to the rest-frame wavelength and requires that troughs extend for at least 2,000 km/s. The presence of BAL of Lyα forest studies is given by spectra with BI > 0.

3.3.3.2 Detection of DLAs

DLAs are usually looked for in spectra with $z \geq 2$ to have a visible Lyα forest. The first fully automated procedure uses Voigt-profile fitting method to detect DLA in quasar spectra. The algorithm searches for strong absorptions in the full line of sight and estimate the absorber redshift and its column density NHI through correlation with synthetic profiles using all data emission lines. A non-public procedure has been applied on DR12 and DR14 quasar spectra.

Accuracy and completeness of the measurements were quantified in Noterdaeme et al. (2012) for DR12. The resulting DLA catalogs give two main purity parameters : cc (correlation at detection), which should be high for high confidence detection, and Pcc (the logarithm of the probability that cc is due to random), which should be low for high confidence detection. Hence, the procedure allows for flexibility in the choice of DLAs depending on the requirements for completness and purity.

The second automated procedure is described in Parks et al. (2018) and uses a convolutional neural network (CNN). The publication also made public the resulting DLA catalog on DR12. Given that the Voigt-profile procedure was not public and had only been run on spectra up to DR14, I led the detection of DLA on DR16 using the CNN procedure, which has been included in the quasar catalog DR16Q (in prep paper). The algorithm searches in the 900 Å $\leq \lambda_{\mathrm{RF}} \leq 1,346$ Å rest-frame window and returns the classification of the identified absorption systems, including DLA identification, DLA location in the spectra and the associated column densities. The CNN was trained with visually-inspected SDSS DLA-free spectra and mock DLA spectra. The training set ignores correlation of metal-line absorptions with DLAs and the clustering of lines with DLAs. They arbitrarily provide a confidence parameter, based on the homogeneity of the predictions for an absorbing system depending on the input portion of the spectra given to the CNN. If high purity is required in the DLA sample, the authors in Parks et al. (2018) advise to take the confidence parameter above 0.9 (1 being the highest value) and absorbing systems with log(NHI) $\geq$ 20.3.

4

Measuring the 1D power spectrum of the Lyα forest with SDSS data

Contents

THE Lyα forest is an insightful probe to the intergalactic neutral gas distribution, which acts as a tracer of matter distribution. In Chapter 2, I showed how efficient is the collection of absorptions seen in the spectra of distant quasars at constraining cosmology. In particular, the 1D power spectrum of the Lyα forest ($P_{\mathrm{Ly}\alpha}$) stands as a powerful tool to estimate the impact of relativistic particles on the evolution of structures on the Mpc scale. However, constructing the $P_{\mathrm{Ly}\alpha}$ from spectroscopic data requires a rigorous and detailed analysis. In this Chapter, I present the measurement of the $P_{\mathrm{Ly}\alpha}$ using data from BOSS and eBOSS of the SDSS-III and SDSS-IV surveys, from a sample of 43,751 high-quality quasar spectra with absorbing redshift $2.1 \leq z_{\mathrm{Ly}\alpha} \leq 4.7$. Such measurements have been explored by several groups before, the first one was by Croft et al. (1998), followed by McDonald et al. (2000), Croft et al. (2002). More recently, the Lyα-Saclay team has become a world-leading group in such analyses, with the measurement of the $P_{\mathrm{Ly}\alpha}$ using about 13,000 quasar spectra from the SDSS-III survey published in Palanque-Delabrouille et al. (2013), to which I will henceforth refer to as PYB13. I took the lead of the most up-to-date measurement of the $P_{\mathrm{Ly}\alpha}$ with more than a factor three improvement on the number of quasars. The required tolerance of the systematic budget must increase accordingly to the statistical gain in order to not biase our scientific interpretations of the data. Therefore, this unprecedented statistical power requires to perform a careful investigation of all the observational systematic errors and their sources in order to provide robust constraints on the cosmological parameters. The outline of this Chapter is as follow. Sec. 4.1 outlines the general scheme for measurement of the $P_{\mathrm{Ly}\alpha}$. The stringent data selection and analysis of the data are detailed in Sec. 4.2. Sec. 4.3 describes the dedicated synthetic data (thereafter mocks) and shows how they are used to quantify and correct for the possible biases introduced in the analysis. The extensive thorough investigation of the systematic uncertainties is outlined in Sec. 4.4. Finally, Sec. 4.5 presents the measured 1D Lyα power spectrum and the resulting improvements compared to PYB13. This analysis has been published in Chabanier et al. (2019) with my PhD advisor Nathalie Palanque-Delabrouille, collaborators at DPhP, Christophe Yèche, Jean-Marc Le Goff, Jim Rich and other collaborators from the eBOSS collaboration.

4.1 Methodology

4.1.1 SDSS data

The results presented in this Chapter are based on data collected by the SDSS York et al. (2000) telescope, presented in Chapter 3. We select our sample of Lyα forest observations from the quasar spectra of the DR14Q catalog Pâris et al. (2018), which were observed either over a five-year period from 2009 to 2014 by the SDSS-III Collaboration Gunn et al. (2006); Ahn et al. (2012); Dawson et al. (2013); Eisenstein et al. (2011); Smee et al. (2013) during the BOSS survey, or in 2014 – 2015 by the SDSS-IV Collaboration Blanton et al. (2017) as part of the eBOSS survey Dawson et al. (2016). The selection of the quasars for either survey is extensively described in Ross et al. (2012); Myers et al. (2015); Palanque-Delabrouille et al. (2016).

While all the quasar targets of the BOSS DR12Q quasar catalog Pâris et al. (2012) underwent visual inspection of the measured spectra, this is no longer the case for eBOSS. This is due to the significant increase in the number of targets, from 40 deg^{-2} quasar targets focusing on redshifts above 2.1 for BOSS, to about 115 deg^{-2} quasar targets for eBOSS, encompassing both quasars in the redshift range $0.8 < z_{\mathrm{qso}} < 2.2$ used as direct matter tracers Ata et al. (2018), and quasars at $z_{\rm qso} > 2.1$ used for Lyα studies. As a consequence, there is a slight increase in the rate of inaccurate redshift determinations, which mostly affects quasars at redshifts above ~ 4 when the Mg II emission line of a low-redshift quasar is mistaken for the Lyα emission line of a high-redshift quasar. This contamination, which is less than 1%, has negligible impact on BAO studies where the bulk of the sample is largely dominated by $z_{\rm qso} \sim 2.5$ quasars, but it is significant for 1D power spectrum analyses where all redshift bins are considered with equal importance. In particular, for $z_{\rm qso} \gtrsim 4$, the fraction of quasars with a wrong assignation of the redshift can reach 30% of the sample. We therefore restrict the sample to quasars observed before MJD=56870 when they were all visually inspected. We use the latest spectral identification, i.e. the one given in DR14Q.

The data are processed with release v5_7_0 based on the standard DR12 SDSS-III pipeline Bolton et al. (2012). We use the coadded spectra resampled at wavelength pixels of width $\Delta \log_{10} \lambda = 10^{-4}$. We also use the individual exposures to improve the estimate of the pixel noise. The noise estimate procedure is described in Sec.4.2.3.

4.1.2 Transmitted flux and quasar continuum

The top plot of Fig. 4.1 displays one eBOSS spectra. The broad quasar emission lines are clearly visible, such as Lyβ (1025.72 Å), Lyα (1215.67 Å), N V (1238.82/1242.80 Å), Si IV (1393.76/ 1402.77 Å) and C IV (1548.20/1550.78 Å), with all wavelengths expressed in rest frame. Lyα absorption along the quasar line of sight, constituting the Lyα forest, appears bluewards of the quasar Lyα emission peak. For illustration purposes, the bottom panel of Fig.. 4.1 shows composite spectra obtained by averaging all quasar spectra used in this analysis, presented later in Sec. 4.2.2, split into six redshift bins. We can clearly see the higher mean absorption (and hence smaller transmitted flux) at higher redshift, due to the larger density of neutral hydrogen as one moves to earlier times.

We define the Lyα forest region by the range $1050\,\text{Å} < \lambda_{\rm RF} < 1180\,\text{Å}$ (colored bands in Fig. 4.1, top panel), about $6000\,{\rm km \cdot s^{-1}}$ and $8500\,{\rm km \cdot s^{-1}}$ from the quasar Lyβ and Lyα emission peaks, respectively, to avoid contamination of the power spectrum by astrophysical effects in the vicinity of the quasar. The Lyα forest region spans a redshift range $\Delta z = (\lambda_{\rm RF,max} - \lambda_{\rm RF,min}) \frac{1+z_{\rm qso}}{\lambda_{\rm Ly\alpha}} \sim 0.4$ for a quasar at a redshift $z_{\rm qso} = 3$, and $\Delta z \sim 0.6$ at $z_{\rm qso} = 4.6$, where $\lambda_{\rm Ly\alpha} = 1215.67\,\text{Å}$. In order to improve the redshift resolution of the measured power spectrum and to reduce the correlation between redshift bins, we split this range into three consecutive and non-overlapping sub-regions of equal length, each covering ~ 170 pixels of eBOSS spectra. The boundaries between these sub-regions are set at rest-frame wavelengths of 1093.3 and 1136.6 Å. For the sake of simplicity, I hereafter use *forest* (and not *a third of forest* or *a sub-forest*, for instance) to refer to each of these sub-regions. Each forest spans at most $\Delta z = 0.2$. The first step of the analysis is to extract these forests from the spectra ; All subsequent steps are applied on each forest.

Following equation (2.15), our estimator for the transmitted flux fraction is $\delta(\lambda)$, is described with the pixel flux, $f(\lambda)$, by

$$\delta(\lambda) = \frac{f(\lambda)}{C_q(\lambda)\overline{F}(z_{\rm Ly\alpha})} - 1\,, \tag{4.1}$$

where $C_q(\lambda)$ is the unabsorbed flux of a quasar (the mean quasar 'continuum') and $\overline{F}(z_{\rm Ly\alpha})$ is the mean transmitted flux fraction at the H I absorber redshift, such that the product $C_q(\lambda)\overline{F}(z_{\rm Ly\alpha})$ represents the mean expected flux. I recall that for a pixel at observed wavelength λ, the

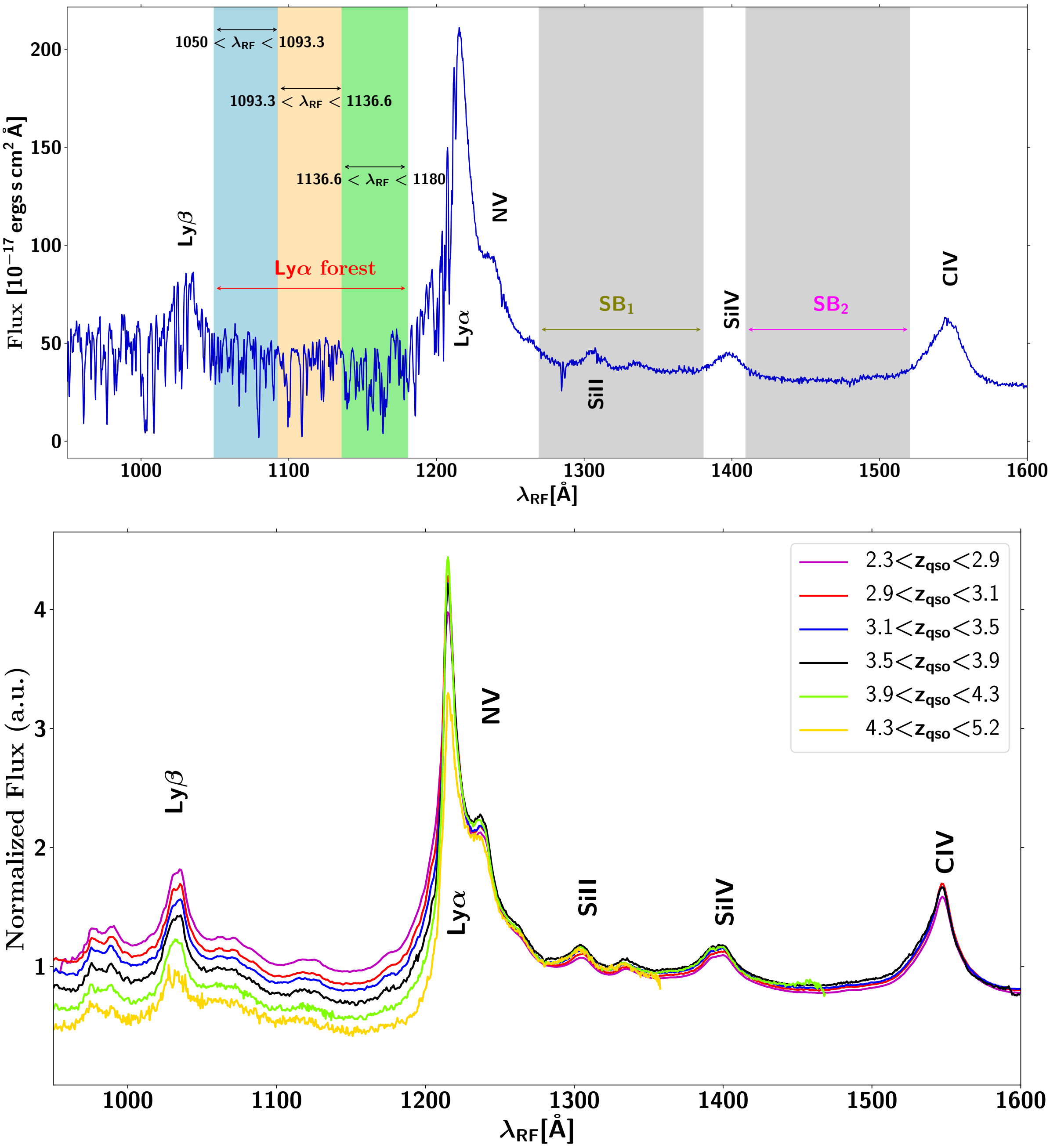

FIGURE 4.1 – *Top : Example of a bright quasar spectrum (SDSS J114308.87+345222.2) at redshift* $z_{qso} = 3.155$*, observed by BOSS. From the left to the right, the three colored bands represent the Lyα forest region (*$1050 < \lambda_{\mathrm{RF}} < 1180$ Å*), while the grey bands show the first (*SB$_1$ *at* $1270 < \lambda_{\mathrm{RF}} < 1380$ Å*) and the second (*SB$_2$ *at* $1410 < \lambda_{\mathrm{RF}} < 1520$ Å*) side bands, in the quasar rest frame. Bottom : Composite quasar spectra in six redshift bins from 2.3 to 5.2. All spectra are normalized at* $\lambda_{RF} = 1280$ *Å.*

corresponding H I absorber redshift $z_{\mathrm{Ly}\alpha}$ can be inferred from equation (2.7). The product $C_q(\lambda)\overline{F}(z_{\mathrm{Ly}\alpha})$ is determined using a method similar to the approach developed in Bautista et al. (2017); du Mas des Bourboux et al. (2017). We assume that the quasar continuum is the product of a universal function of the rest-frame wavelength, $\lambda_{RF} = \lambda/(1+z_{\mathrm{qso}})$, and a quasar-dependent factor a_q :

$$C_q(\lambda) = a_q C(\lambda_{RF}) \;, \tag{4.2}$$

where $C(\lambda_{RF})$ assumes no functional form, and is normalized so that its integral over the forest is equal to unity. The a_q and $C(\lambda_{RF})$ are determined iteratively by maximizing the likelihood function

$$\mathcal{L} = \prod_{q,\lambda} P(\, f(\lambda) \mid C_q(\lambda)\,)\;. \tag{4.3}$$

Here $P(f(\lambda) \mid C_q(\lambda))$ is the probability to observe a flux $f(\lambda)$ with the observational resolution assumed to be Gaussian.

The method uses the same publicly available pipeline `picca`[1] (Package for IGM Cosmological-Correlations Analyses). In contrast to BAO analyses where pixels are weighted by the noise inverse variance, we here do not any apply weights. All pixels of a given quasar thus contribute equally to the measurement of $C_q(\lambda)\overline{F}(z_{\mathrm{Ly}\alpha})$, just as they will contribute equally to the FFT computation described in the next section.

4.1.3 Power spectrum estimation

The fluctuations of the transmitted flux around its mean value can be decomposed in three terms,

$$\delta(\lambda) = \delta_{\mathrm{Ly}\alpha}(\lambda) + \delta_{\mathrm{background}}(\lambda) + \delta_{\mathrm{noise}}(\lambda), \tag{4.4}$$

where $\delta_{\mathrm{Ly}\alpha}$ and $\delta_{\mathrm{background}}$ are caused by Lyα neutral hydrogen absorptions and other metals absorptions respectively, and δ_{noise} is the noise fluctuation uncorrelated to any physical absorptions. $\delta_{\mathrm{background}}$ can be further developed into $\delta_{\mathrm{Si\,III/II}} + \delta_{\mathrm{metals}}$ where $\delta_{\mathrm{Si\,III/II}}$ is the absorption of either Si III or Si II correlated to Lyα absorptions with wavelength separation $\Delta\lambda = 9.2\,$Å, and δ_{metals} is the absorption of uncorrelated metals, i.e. absorptions with very large wavelength separations, such as C IV or Si IV. Then we take into account the spectral response of the spectrograph by convolving the signal with a top-hat function Π for the resolution and a sampling function Ш for the pixellisation, giving,

$$\delta(\lambda) = \Big(\delta_{\mathrm{Ly}\alpha}(\lambda) + \delta_{\mathrm{Si\,III/II}}(\lambda) + \delta_{\mathrm{metals}}(\lambda)\Big) \circledast \Pi(\lambda) \circledast \text{Ш}(\lambda) + \delta_{\mathrm{noise}}(\lambda). \tag{4.5}$$

The noise is not affected by instrumental effects because it is related to number of photons hitting CCD pixels.

Our estimator of the 1D Lyα power spectrum of the Lyα forest ($P_{\mathrm{Ly}\alpha}$) is P_{1D}, and we use the intermediate 1D power spectrum of the flux transmission field P_{raw}. To measure P_{raw}, we decompose each absorption spectrum $\delta(\lambda)$ into Fourier modes and estimate their variance as a function of wave number k. In practice, we do this by computing the discrete Fourier transform using a fast Fourier Transform (FFT) algorithm of the δ field. The use of a FFT requires that the pixels be equally spaced. This condition is satisfied with the quasar coadded spectra provided by the SDSS pipeline (Bolton et al., 2012) with a constant pixel width $\Delta[\log(\lambda)] = 10^{-4}$ and the

---

[1] `https://github.com/igmhub/picca`

velocity difference between pixels, i.e. the relative velocity of absorption systems at wavelength $\lambda + \Delta\lambda/2$ and $\lambda - \Delta\lambda/2$ is

$$\Delta v_{\rm pix} = c\frac{\Delta\lambda}{\lambda} = c\Delta[\ln(\lambda)] = c\ln(10)\Delta[\log(\lambda)] = 69\ \mathrm{km}\cdot\mathrm{s}^{-1}. \tag{4.6}$$

We measure the power on scales small enough for this expression to be valid. Throughout this analysis we therefore use velocity instead of observed wavelength. Similarly the wave vector $k \equiv 2\pi/\Delta v$ is measured in $\mathrm{s}\cdot\mathrm{km}^{-1}$. The highest k-mode possible is determined by the Nyqvist-Shannon limit at $k_{\rm Nyqvist} = \pi/\Delta v_{\rm pix}$, i.e. $k_{\rm Nyqvist} = 0.045(\mathrm{km}\cdot\mathrm{s})^{-1}$. We limit the analysis, however, to a maximal mode $k_{\rm max} = 0.02\ \mathrm{s}\cdot\mathrm{km}^{-1}$, beyond which the spectrograph resolution cuts the power by over a factor of ten. The smallest k-mode possible is limited by the lenght of the spectra, i.e. by the redshift resolution $\sim \Delta z = 0.2$. We express each absorption spectrum in velocity units $\delta(\Delta v)$ and take the square of the Fourier transform $P_{\rm raw}(k) = |\mathcal{F}(\delta(\Delta v))|^2$. For each spectra we have

$$P_{\rm raw}(k) = (P_{\rm 1D}(k) + P_{\rm metals}) \cdot W^2(k, R, \Delta v) + P_{\rm noise}(k). \tag{4.7}$$

where $W^2(k, R, \Delta v)$ is the window function corresponding to the spectral response of the spectrograph. The window function depends on the pixel width and on the spectrograph resolution R, such that

$$W(k, R, \Delta v) = \exp\left(-\frac{1}{2}(kR)^2\right) \times \frac{\sin(k\Delta v/2)}{(k\Delta v/2)}\,. \tag{4.8}$$

The resolution is derived from measurements obtained with spectral lamps, as described in Smee et al. (2013), and is provided by the eBOSS reduction pipeline Bolton et al. (2012) for each coadded spectrum. Since the measurement of the 1D power spectrum on small scales is extremely sensitive to the resolution, we adopt in this analysis the approach extensively described in PYB13. This previous work provides a table of corrections of R as a function of the position on the CCD in terms of fiber number and wavelength. As in PYB13, we apply the correction to each pixel of each spectrum. Fig. 4.2 illustrates the window function in function of different resolution values R.

The $P_{\rm 1D}$ term in the equation (4.7) is our estimator for $P_{\rm Ly\alpha}$ in this analysis. However, I stress that it does not correspond exactly to the power spectrum of the Lyα forest because, in addition of the Lyα signal, its contains the correlated absorption of Lyα and Si III or Si II within the Lyα forest and can be estimated directly in the power spectrum. Since Si III absorbs at $\lambda_{\rm RF} = 1206.50$ Å, just 9 Å from Lyα, its absorption appears in the power spectrum as oscillations with a frequency corresponding to $\Delta v_{\rm Si\,III} \sim 2271\,\mathrm{km}\cdot\mathrm{s}^{-1}$. Its contribution cannot be isolated from the Lyα absorption at this level of the data anlysis. Similarly, absorption by Si II at $\lambda_{\rm RF} = 1190$ and 1193 Å creates an oscillatory pattern at a frequency corresponding to $\Delta v_{\rm Si\,II} \sim 5577\,\mathrm{km\ s}^{-1}$. We perform a theoretical modeling of this oscillatory feature in order to only take into account the signal caused by Lyα absorption at the cosmological analysis step. This will be described in Chapter 7. Other metal absorptions occur at large wavelength separation, then because of the negligible correlations we can isolate them in the data analysis. The other terms of equation (4.7), the noise power spectrum $P_{\rm noise}(k)$ and the metal power spectrum $P_{\rm metals}(k)$, undergo specific updated treatments compared to the analysis of PYB13, and are described in Sec.s 4.2.3 and 4.2.4.

We compute the Fourier transform using the efficient FFTW package. The mean redshift of the Lyα absorbers in a forest determines the redshift bin to which the forest contributes. We

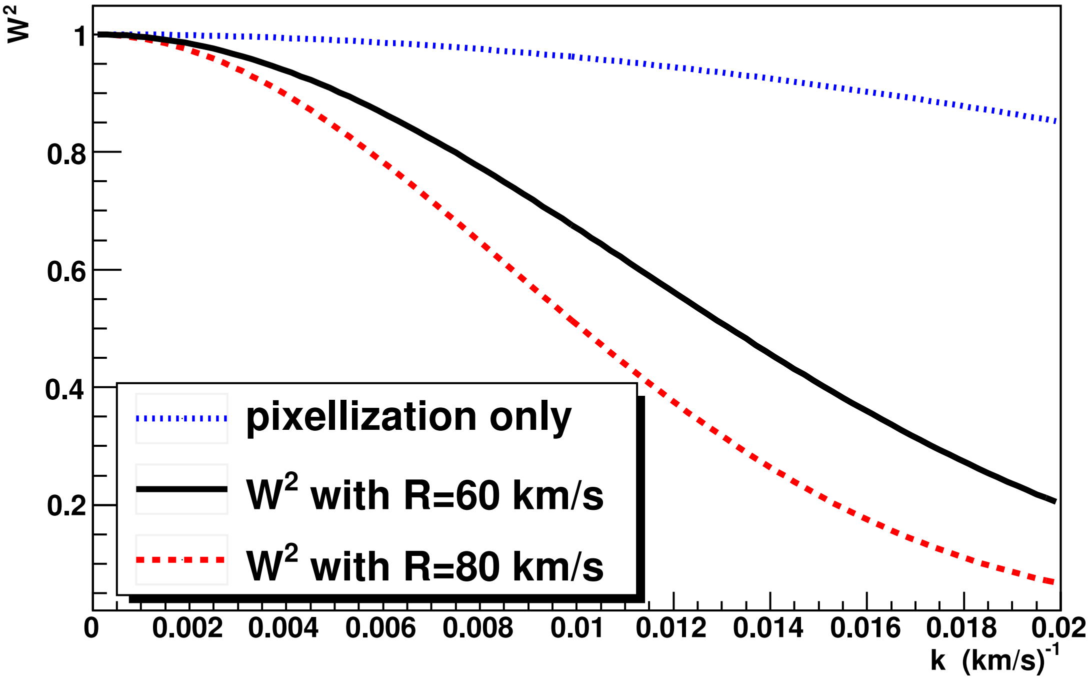

FIGURE 4.2 – *Window function* $W^2(k, \bar{R}, \Delta v)$*, with* $\Delta v = 69\,\mathrm{km\cdot s^{-1}}$*, reproducing the spectrum binning and the impact of the spectrograph resolution, for a resolution* $\bar{R} = 60\,\mathrm{km\cdot s^{-1}}$ *typical at* $\lambda > 5000\,\mathring{A}$ *and* $\bar{R} = 80\,\mathrm{km\cdot s^{-1}}$ *typical at* $\lambda < 4300\,\mathring{A}$*. For comparison, we also show the contribution from the pixellization only (equivalent to* $\bar{R} = 0$*).*

rebin the final power spectrum onto an evenly spaced grid in k-space, assigning equal weight to the different Fourier modes that enter each bin. Using equation (4.7), the final 1D power spectrum, $P_{\rm 1D}$, which combines the correlation in Lyα absorptions and the correlation between neutral hydrogen and Si III or Si II, is obtained by averaging the corrected power spectra of all contributing forests, as expressed in the following estimator :

$$P_{\rm 1D}(k) = \left\langle \frac{P_{\rm raw}(k) - P_{\rm noise}(k)}{W^2(k, R, \Delta v)} \right\rangle - P_{\rm SB_1}(k), \tag{4.9}$$

where $\langle\rangle$ denotes the ensemble average over forest spectra and where $P_{\rm SB_1}(k)$ is the power spectrum measured in the first side band corresponding to the wavelength range $1270 < \lambda_{\rm RF} < 1380\,\mathring{\rm A}$ as shown in Fig. 4.1. In equation (4.9), $P_{\rm SB_1}$ includes the power from uncorrelated metals $P_{\rm metals}$ but also a contaminating contribution coming from the spectroscopic pipeline. Details are given in Sec. 4.2.4. The final result is presented in 35 evenly spaced k-modes with $\Delta k = 5.4 \times 10^{-2}\,\mathrm{s\cdot km^{-1}}$, in 13 evenly spaced redshift bins from $z_{\rm Ly\alpha} = 2.2$ to 4.6.

A second approach to estimate the 1D flux power spectrum was also used in PYB13 based on a likelihood method to compute the covariance matrix of the δ field in real space as a function of the pixel-pair separation in the spectra (McDonald et al., 2006). It is more precise than the FFT method because it can take variation in the noise or in the spectrograph resolution at the pixel level and offers a natural way to mask pixels affected by sky emission lines or DLAs, which will be explained in the Sec. 4.3.3 for the FFT method. However, the likelihood method can be subject to convergence problems in the presence of noisy spectra and is more sensitive to the implementation of the analysis method, such as for the quasar continuum estimation. The comparison in PYB13 demonstrated that the two methods yield compatible results, and we choose the FFT approach given that it is much faster.

4.2 Data analysis

4.2.1 Value added catalogs

In addition to the SDSS quasar catalogs, we make use of additional catalogs and information which help in the selection of the spectra for the analysis. For instance, we want to exclude spectra exhibiting DLAs or quasars affected by BALs, since these features are not included in the simulations we use for cosmological parameter inferences. We use the Balnicity Index available in the DR14Q catalog as described in Sec. 3.3.3, which flags quasars with BALs in their spectra. Finally, we examine two external catalogs to identify regions in the quasar spectra affected by DLA systems resulting from the automated methods described in Sec. 3.3.3. The first is an update on DR14Q of the identification of DLAs following the Voigt-profile line fitting procedure. We will hereafter refer to it as N12. The second catalog is constructed from the CNN procedure. We will refer to the latter catalog as P18. Details on the use of these additional catalogs for the forest selection are given in Sec. 4.2.2, and for the estimate of the associated systematic uncertainties in Sec. 4.3.1. We also use a list of wavelength regions contaminated by sky emission lines. This list is available at `https://github.com/igmhub/picca/blob/master/etc/list_veto_line_Pk1D.txt`

4.2.2 Quasar selection

The DR14Q quasar catalog contains 525,982 quasars. We are solely interested in the 209,407 quasars observed by BOSS or eBOSS at $z_{\mathrm{qso}} > 2.1$ for which the Lyα region is accessible. When restricting to visually inspected spectra (those with MJD < 56870), the sample reduces to 180,413 objects. Finally, we discarded quasars with BAL features, as flagged by a non-zero value of BI_CIV, bringing the initial sample to 167,988 quasars. In case of multiple observations of a quasar, we only use the best one.

The analysis of PYB13 was reaching a similar level of statistical and systematic uncertainties on some modes or redshifts, despite being based upon an initial sample of only about 60,000 $z_{\mathrm{qso}} > 2.1$ quasars, from which about $14,000$ were selected. The significant increase in statistics of the present sample compared to PYB13 makes it possible, if not mandatory, to tighten the selection criteria in order to reduce the impact of the systematic uncertainties on the measured power spectrum. The following criteria are applied on each forest, and no longer on the selection of the quasar spectra.

To improve the quality of the low-redshift forests, we increased the CCD short-wavelength cut, below which the CCD becomes considerably noisier, from 3650 to 3750 Å. We discard forests where the mean spectral resolution is larger than $85\,\mathrm{km}\cdot\mathrm{s}^{-1}$. We also optimize the threshold on the mean signal-to-noise ratio per pixel (SNR) below which we reject a forest, where the SNR is defined as the ratio of the pixel flux to the pixel noise, and the average is computed after pixel masking. This optimization is done by computing, for each redshift bin, the uncertainty on the mean value of P_{1D}, as a function of the threshold on the mean SNR per pixel. This uncertainty depends upon the range of modes considered, as illustrated in Fig. 4.3 where we test the P_{1D} uncertainty in samples of varying minimum SNR. On large scales where signal dominates, the uncertainty decreases with more statistics from a more inclusive sample, and hence a lower threshold, whereas on small scales where noise tends to dominate, the higher the threshold the lower the uncertainty. We therefore set the SNR threshold so as to minimize the uncertainty around a central mode, at $k \sim 0.01\,\mathrm{s\cdot km^{-1}}$. We checked that our final power spectrum measurement does not contain significant variations with a ±0.3 shift in the chosen SNR threshold. For $z_{\mathrm{Ly}\alpha} \geq 3.4$, the uncertainty no longer evolves significantly with SNR, and the threshold is fairly independent of redshift. As a result, we apply the thresholds given in Tab. 4.1 for the forest selection. For comparison, it was set to 2.0 for all redshifts in PYB13.

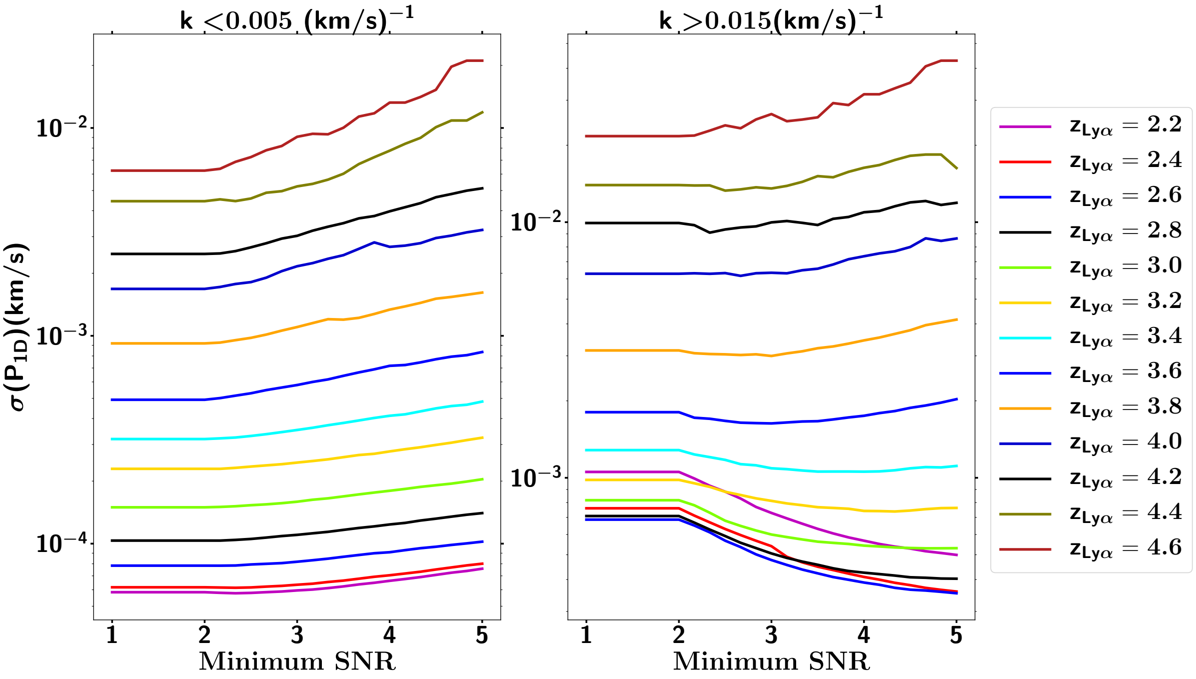

FIGURE 4.3 – *Uncertainty on the mean value of P_{1D} for the 13 redshift bins, focusing on the average over either large scales ($k < 0.005$, left panel) or small scales ($k > 0.015$, right panel).*

$\langle z_{\mathrm{Ly}\alpha}\rangle$	2.2	2.4	2.6	2.8	3.0	3.2	3.4	3.6 to 4.6
SNR threshold	4.1	3.9	3.6	3.2	2.9	2.6	2.2	2.0

TABLE 4.1 – *Threshold on the minimum mean signal-to-noise ratio per pixel in a forest*

Sky lines affect the data quality by increasing the pixel noise. We mask major sky lines (e.g., lines at 5577 Å, 5890 Å, 6300 Å, 6364 Å, 6864 Å) in each forest. We locate the position of DLAs in our forests using catalog N12 introduced in Sec. 4.2.1. We mask the DLAs following the same procedure as in Bautista et al. (2017) : all pixels where the DLA absorption is higher than 20% are masked, and the absorption in the wings is corrected using a Voigt profile.

We discard forests shorter than 75 pixels, where the reduced length can be due to the CCD UV cut or to the presence of a strong DLA at the forest boundary. We also discard forests with more than 40 masked pixels, whether from DLA absorption, sky line masking, or flags from the SDSS pipeline (the latter are indicated by a null variance). Since the use of a FFT to compute the power spectrum requires equally-spaced pixels, we reintroduce all masked pixels in the forest before performing the Fourier transform and set their flux to zero. This procedure introduces a

k-dependent bias in the resulting power spectrum, which we quantify and correct as discussed in Sec. 4.3.3.

This tight selection procedure yields a sample a 43,751 high-quality quasar spectra, from which we use 94,558 forests[2] for the analysis. The yields per redshift bin are summarized in Tab. 4.2. The UV cut, the resolution cut, and the stringent SNR cut at low redshift contribute to often discarding the first (or even first two) forests of a given quasar spectrum, since the blue end of the spectrograph suffers from all these drawbacks (large noise and poor resolution). These cuts explain why the number of forests is not simply three times the number of quasars. Fig. 4.4 presents the resulting redshift and resolution distribution of our selected forests.

z bin	2.2	2.4	2.6	2.8	3.0	3.2	3.4
#	17,144	20,089	16,541	14,762	10,364	6,767	4,763
$\langle z_{\mathrm{Ly}\alpha} \rangle$	2.207	2.396	2.595	2.795	2.991	3.190	3.393
⟨SNR⟩	7.3	7.6	7.3	6.8	6.5	6.0	5.2
⟨R⟩	81.2	77.8	73.9	71.0	68.9	67.3	66.1
z bin	3.6	3.8	4.0	4.2	4.4	4.6	
#	2,356	933	421	229	126	63	
$\langle z_{\mathrm{Ly}\alpha} \rangle$	3.587	3.786	3.994	4.194	4.387	4.578	
⟨SNR⟩	4.6	4.3	3.9	4.0	3.8	3.5	
⟨R⟩	65.2	65.9	70.6	72.9	71.2	69.3	

TABLE 4.2 – *Summary per redshift bin of the number of forests, the mean redshift, the mean SNR per pixel, and the mean resolution.*

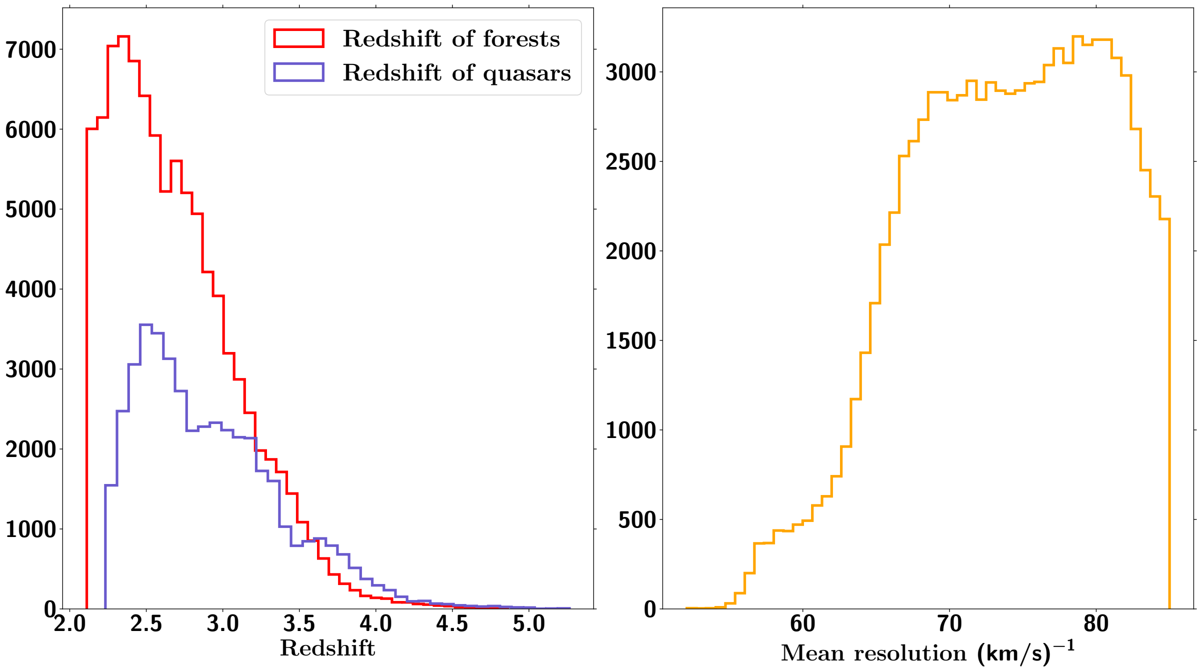

FIGURE 4.4 – *Distributions of mean forest and mean quasar redshift on the left, and mean resolution per pixel on the right for the selected forest sample.*

---

[2] Recall that *forests* are sub-regions defined in Sec. 4.1.2 of the commonly-called Lyα forest of quasar spectra.

4.2.3 Estimator of noise power

The accuracy with which noise is determined in the automated pipeline, as described in Chapter 3, is insufficient for the purpose of this analyse given the level of precision allowed by the increasing statistics. A technique has been developped in McDonald et al. (2006), also used in PYB13, to construct an appropriate estimate.This refinement is not necessary for measuring the large-scale 3D autocorrelation function because the noise in the data only affects the amplitude of the power spectrum. Because we aim at measuring the absolute level of the power spectrum, which is sensitive to the suppression induced by relativistic particles such as active or sterile neutrinos, we are directly affected by the level of noise and it is crucial to properly account for it. We use the $N_{\rm exp}$ individual exposures used to construct the coadd quasar spectrum to compute the noise power spectrum $P_{\rm noise}$ in a two-step process.

We first construct a spectrum that contains the same noise as the data coadded spectrum, but devoid of any power from an astrophysics or cosmology signal. To this end, we compute the semi-difference $\Delta\phi$ between two customized coadded spectra obtained from weighted averages of the even-number exposures, for the first spectrum, and of the odd-number exposures, for the second one, with weights taken as the pixel inverse variances. In this first step, we assign zero weight to pixels flagged by the COMBINEREJ bit of the SDSS pipeline, i.e., pixels rejected when computing the coadded spectrum from the individual exposures. Even assuming all exposures to have the same variance, the variance measured in $\Delta\phi$ is the same as the variance in the data coadded spectrum only if $N_{\rm exp}$ is even. In contrast, if $N_{\rm exp}$ is odd, we can write $N_{\rm eff} = 2N_1 + 2$, and the variance of $\Delta\phi$ is reduced by $[1/N_1 + 1/(N_1 + 1)]/4$ with respect to the variance in each exposure, instead of by the $1/N_{\rm exp}$ factor that is expected for the data coadded spectrum. To account for a possible difference in the number of exposures of the two customized spectra, we thus renormalize $\Delta\phi$ by the factor $\sqrt{\alpha}$, where $\alpha = 4 \times E[N_{\rm exp}/2)] \times E[(N_{\rm exp} + 1)/2)]/N^2_{\rm exp}$, with E denoting the truncated integral part. This approach gives $\alpha = 1$ if $N_{\rm exp}$ is even, and, for instance, $\alpha = 0.98$ for a quasar with five exposures. The quantity $P^{\rm diff}_{\rm noise}(k) = |\mathcal{F}(\alpha\Delta\phi)|^2$, where $\mathcal{F}(\alpha\Delta\phi)$ is the Fourier transform of the normalized difference spectrum, is expected to be an accurate estimator of the noise power in the coadded spectrum. $P^{\rm diff}_{\rm noise}(k)$ is found to be scale-independent to an accuracy sufficient for our purpose, as expected for a white noise. We thus define $P^{\rm diff}_{\rm noise}$ as the average of $P^{\rm diff}_{\rm noise}(k)$ over all k-modes covered by our analysis.

Because the spectrograph window function W^2 quickly drops to zero beyond the scales used in the analysis, we can further refine the estimate of the noise power spectrum by checking that on very small scales, $P^{\rm diff}_{\rm noise}$ is indeed a lower asymptote of the raw power spectrum $P_{\rm raw}$, as expected from equation (4.7). Fig. 4.5 shows that this is not the case at small redshifts, due to subtle differences between the pipeline coaddition and the above semi-difference computation. In a second step, for each of the 13 redshift bins, we therefore fit $P_{\rm raw}$ on k-modes above $k_{\rm max} = 0.02\,{\rm s\cdot km^{-1}}$ by an exponential decrease plus a constant $P^{\rm lim}_{\rm raw}$ (shown as the blue dashed line in Fig. 4.5). We compute the ratio β between $P^{\rm diff}_{\rm noise}$ and $P^{raw}_{\rm lim}$. We define $P_{\rm noise} = P^{\rm diff}_{\rm noise}$ for redshift bins where $\beta < 1$, and $P_{\rm noise} = P^{\rm diff}_{\rm noise}/\beta$ otherwise. We find β of order 0.95 for the first three redshift bins, and we set it to 1.0 for higher redshifts.

4.2.4 Estimation of uncorrelated background power spectrum

The metal power spectrum $P_{\rm metals}(k)$ of equation (4.7), corresponding to uncorrelated background due to metal absorption in the Lyα forest, is independent of Lyα absorption and cannot be estimated directly from the power spectrum measured in the Lyα forest because the wave-

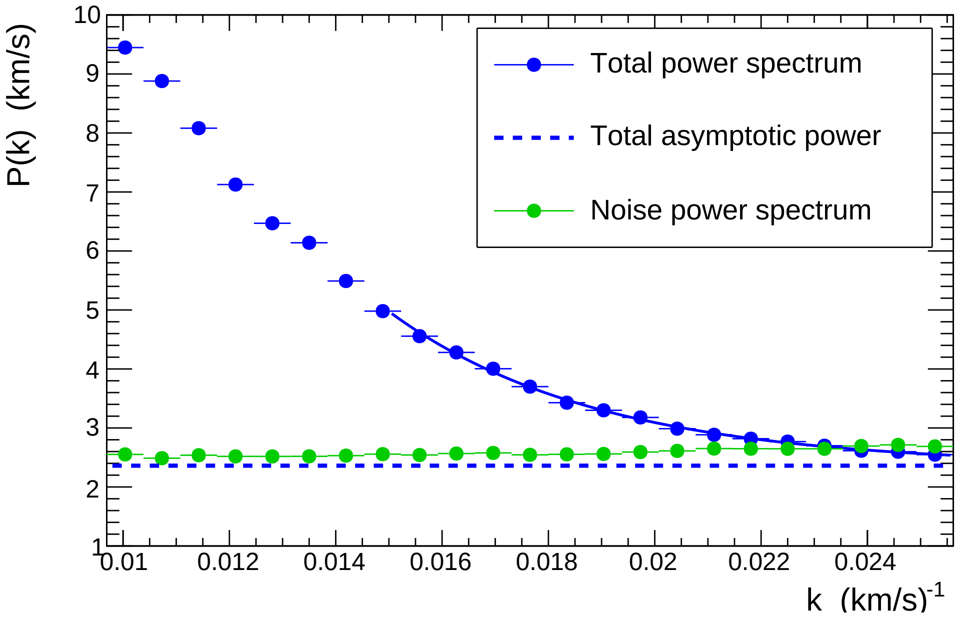

FIGURE 4.5 – *Power spectra of the noise (green circles), raw data (blue circles), and the asymptote of the total power spectrum (dashed blue line), for $z < 2.5$. A correction is applied to the noise level when the latter exceeds the raw power spectrum.*

length separation is too large. The authors in McDonald et al. (2006) and PYB13 addressed this issue by estimating the power spectrum in side bands located at longer wavelengths than the Lyα forest region, and the power spectrum was subtracted from the Lyα power spectrum measured in the same gas redshift range. This method presents the advantage of allowing us to simultaneously account for most residual effects affecting our determination of the 1D power spectrum, we will use the same in this analysis.

We define two side bands, SB$_1$ and SB$_2$, corresponding, respectively, to the wavelength range $1270 < \lambda_{\mathrm{RF,SB1}} < 1380$ Å and $1410 < \lambda_{\mathrm{RF,SB2}} < 1520$ Å as shown in Fig. 4.1. The power spectrum measured in the first side band, SB$_1$, contains the complete contribution from all metals with $\lambda_{\mathrm{RF}} > 1380$ Å, including in particular absorption from Si IV and C IV. The second side band, SB$_2$ also includes C IV, but not the Si IV absorption. We thus use SB$_1$ to subtract the metal contribution in the power spectrum, and SB$_2$ as an important consistency check.

Our method is purely statistical : for a given redshift bin, we use different quasars to compute the Lyα forest and the metal power spectra. For instance, the first redshift bin, $2.1 < z_{\mathrm{Ly}\alpha} < 2.3$ corresponds to Lyα absorption in the observed wavelength range $3750 < \lambda < 4011$ Å from equation (2.7). The contribution of metal absorptions from SB$_1$ in this observed wavelength range is estimated with quasars with a redshift $z_{\mathrm{qso}} \sim 1.9$ using $z_{\mathrm{qso}} = \frac{\lambda}{\lambda_{\mathrm{RF,SB1}}}$. Quasars in a given redshift window have their two side-bands corresponding to fixed observed wavelength windows, which in turn match a specific redshift window of Lyα forest. This extraction of metal power is not an approximation because we are substracting the metal power in the same observed wavelength range as the Lyα forest power from which it is being substracted. The same gas, at the same redshift is doing the absorption both inside the forest and outside in the side-bands, so the absorption will have the same statistical properties.

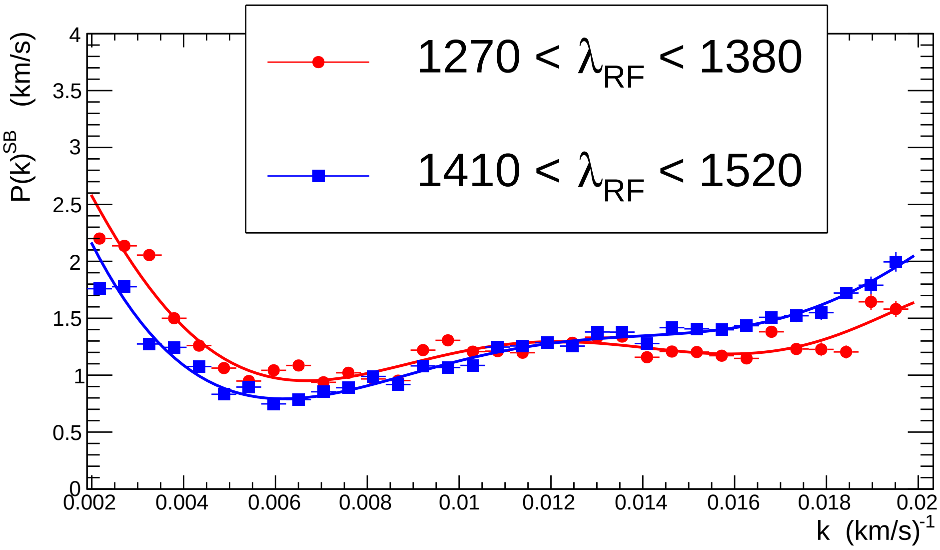

FIGURE 4.6 – *Power spectrum $P_{\mathrm{SB}}(k)$ computed for side band regions on the red side of the Lyα forest. The red dots and the blue squares represent the two side bands defined by* $1270 < \lambda_{\mathrm{RF}} < 1380$ Å *and* $1410 < \lambda_{\mathrm{RF}} < 1520$ Å*, respectively. Each power spectrum is fitted with a sixth-degree polynomial.*

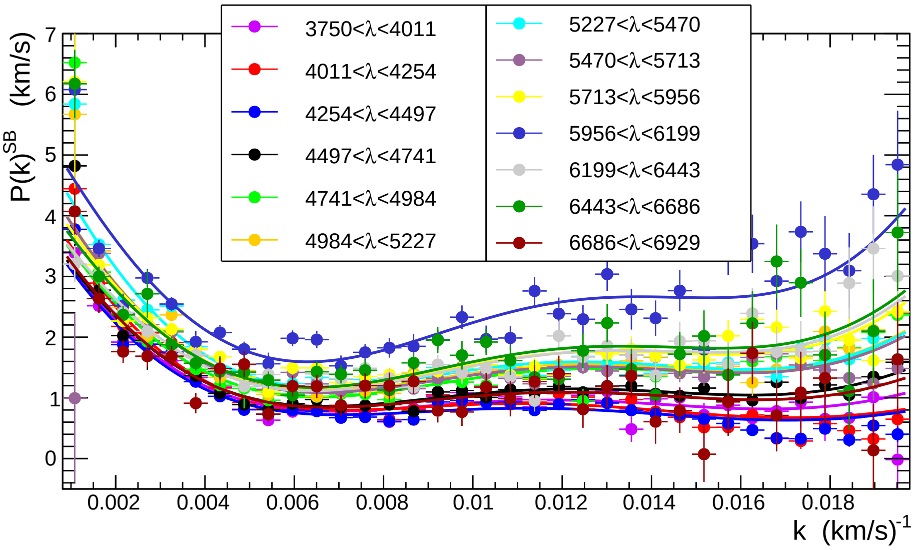

FIGURE 4.7 – *Power spectrum $P_{\mathrm{SB}_1}(k)$ computed for the first side band region* $1270 < \lambda_{\mathrm{RF}} < 1380$ Å *redward of the Lyα forest for the 13 different λ windows. Each λ region corresponds to one redshift bin. Each power spectrum is fitted by the product of the sixth-degree polynomial obtained in Fig. 4.6 and a first-degree polynomial in which the two parameters are free.*

The power spectra $P_{\rm SB}(k)$ shown in Fig. 4.6 are obtained, respectively, for $\rm SB_1$ and $\rm SB_2$ with about 115,000 and 140,000 quasars passing similar quality cuts as the quasars selected for the Lyα forest analysis. The shapes of $P_{\rm SB}(k)$ are similar for the two side bands. As expected, for $\rm SB_2$, which excludes Si IV, the magnitude of $P_{\rm SB}(k)$ is smaller for $k < 0.01 \rm s \cdot km^{-1}$. On smaller scales, the shape is dominated by residuals effects and systematics that exceed the strength of the metal absorption. On small scales, SB2 exhibits more power than SB1 because the measurement includes more quasar spectra affected by the excess of power in the $\lambda \sim 6000\,\text{Å}$ region, as explained below. We fit the distribution $P_{\rm SB}(k)$ with a sixth-degree polynomial. We will use this fitted function as a template to parametrize the $P_{\rm SB}(k)$ measured for each wavelength window (see Fig. 4.6).

As the shape and the magnitude of the power spectrum vary from one wavelength window to another, we parameterize $P_{\rm SB}(k)$ as the product of the fixed shape of Fig. 4.6 by a first-degree polynomial with two free parameters that differ for each wavelength window. This model adequately fits the measured power in all the wavelength windows (see Fig. 4.7). From these parametric functions, we extract the value of the power spectrum $P_{\rm SB_1}(k)$ for each k and for each Lyα redshift window.

As mentioned above, Fig. 4.7 suggests that $P_{\rm SB_1}(k)$ does not decrease to zero at small scales, unlike what would have been expected due to thermal broadening, just as for the Lyα power spectrum. This effect is more pronounced for the three redshift bins that contain the overlap between the two arms of the spectrograph, $5800 < \lambda < 6350\,\text{Å}$, indicating that the excess is likely due to imperfections in the co-addition between the two parts of the spectrum. This behavior also occurs for side bands as well as for Lyα. Therefore, subtracting the power spectrum computed in the side bands not only removes the contribution due to metal absorption but also corrects for residual effects of the pipeline. In the most dramatic case ($\lambda \sim 6000\,\text{Å}$), the residual effect measured at high k corresponds only to $\sim 10\%$ of the power spectrum measured in the Lyα forest. This correction leads to a 3% systematic uncertainty (cf. Sec. 4.4), which is small compared to the statistical error bar.

4.3 Synthetic data and bias corrections

In this section, we investigate the biases introduced at each step of the data analysis, and estimate their impact using mock spectra. They arise from the estimated value of $C_q(\lambda)\overline{F}(z_{\rm Ly\alpha})$ and from the masking of pixels affected by sky emission lines or absorption by DLAs. We use independent multiplicative corrections. Using multiplicative or additive correction would have the same effect given that the amplitude of the corrections are small, and independent because the the continuum estimation and the masking of pixels are uncorrelated at first order.

4.3.1 Mocks

To test the analysis procedure and investigate systematic errors, we generated mock spectra that reproduce the essential physical and instrumental characteristics of the eBOSS spectra. The mocks are produced following the procedure described below. First, a redshift and a g-magnitude are chosen at random from distributions tuned to data. Second, an unabsorbed flux spectrum is drawn for each quasar from a random selection of PCA amplitudes following the procedure of Paris et al. (2011), and the flux normalized to the selected g magnitude. Third, the Lyα forest absorption is generated following a procedure adapted from Font-Ribera et al. (2012), who provide an algorithm for generating any spectrum of the transmitted flux fraction $F(\lambda)$ from a Gaussian random field $g(\lambda)$. Specifically, they present a recipe for choosing the parameters a and b and the power spectrum $P_g(k)$ such that the transformation $F(\lambda) = \exp[-a \exp(bg(\lambda))]$ yields the desired power spectrum and mean value of $F(\lambda)$. In practice we generate a suite of transmitted-flux-fraction spectra for thirteen redshifts that reproduce the observed power. For each wavelength pixel, $F(\lambda)$ is obtained by interpolation between redshifts according to the actual Lyα absorption redshift of the pixel. The unabsorbed flux is multiplied by $F(\lambda)$ and convolved with the spectrograph resolution. The spectra are generated with a pixel width that is three times small than a SDSS pixel, and about three times smaller than the SDSS spectral resolution. We checked that this size was small enough to properly take into account the spectral resolution. Finally, noise is added according to eBOSS throughput and sky noise measurements as was done in Le Goff et al. (2011), and the spectrum is rebinned to the SDSS format. We generated two sets of mock spectra. In the first one, the quasars distribution is the same as the one resulting from our sample selection described in Sec. 4.2.2. The second set contains twenty times the quasar distribution of redshift z > 3.7 to improve statistical uncertainty in the high-redshift bins.

4.3.2 Continuum estimation effect

As a starting point, we checked that computing $\delta(\lambda)$ of equation (4.1) with the generated values of the quasar continuum, $C_q(\lambda)$, and of the mean transmitted flux, $\overline{F}(z_{\mathrm{Ly}\alpha})$, allows an accurate reconstruction of the input power in the absence of noise and resolution effects. This step validates the implementation of the code that computes the power spectrum of a δ field.

We then use the value of $C_q(\lambda)\overline{F}(z_{\mathrm{Ly}\alpha})$ that we estimate as explained is Sec. 4.1.2. We define the bias induced by the continuum estimation as the ratio of the measured flux power spectrum, $P_{measured}$, to the flux power spectrum that was generated in the mock spectra, P_{input},

$$b(k,z) = \frac{P_{measured}(k,z)}{P_{input}(k,z)}. \tag{4.10}$$

The mocks were tuned so that $P_{input}(k, z_{\mathrm{Ly}\alpha})$ and $\overline{F}(z_{\mathrm{Ly}\alpha})$ reproduce the data values. Fig. 4.8 illustrates the measured bias.

The use of equation (4.2) for the continuum instead of the real quasar continuum introduces correlated noise in the estimate of $\delta(\lambda)$, which results in a bias larger than unity. The shape of this bias as a function of k depends on the relative amplitudes of the LSS power spectrum and that of this correlated noise, and therefore evolves with redshift.

We also tested a quasar-dependent term of the form $a_q + b_q(\lambda_{RF} - \overline{\lambda}_{RF})$ where $\overline{\lambda}_{RF}$ is the mean over the forest. We validate the method principle by verifying that whatever the form used for the quasar-dependent term, the power spectra of the data after correction by the relevant correction functions (via equation (4.10)) yield consistent results, with less than 1% difference between the two options. The latter form yields a larger bias than the simpler multiplicative constant, so we do not consider it further.

4.3.3 Pixel masking effect

We mask pixels affected by strong absorption caused by DLA or by emission from sky lines, by setting their flux to zero. Otherwise, sky lines would impact the data quality by increasing significantly the pixel noise, and DLAs, responsible for saturated absorptions on the scale of

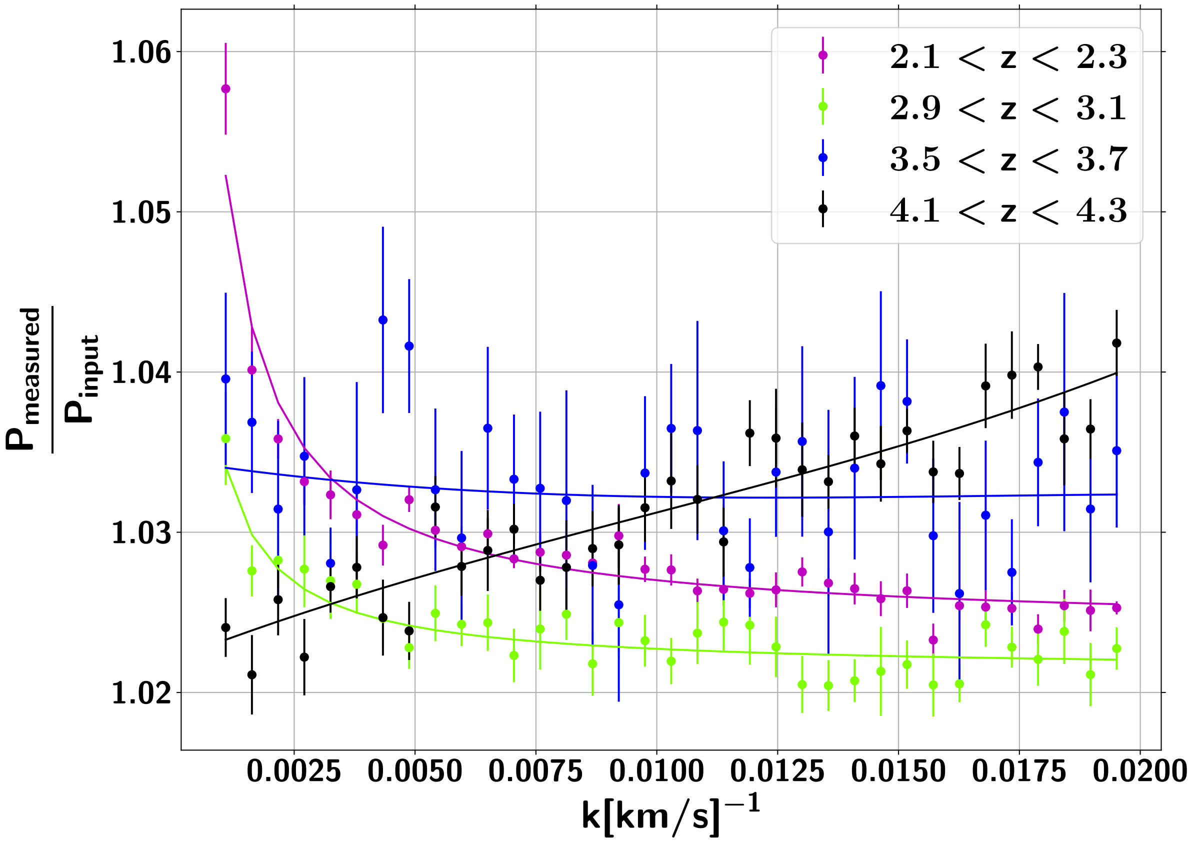

FIGURE 4.8 – *Ratio of measured power spectrum* $P_{measured}$ *to the flux spectrum that was generated in the mock spectra* P_{input}, *illustrated for four redshift bins. The correction is fitted with a functional* $a/k + b$ *dependence for* $2.1 < z_{\mathrm{Ly}\alpha} < 3.5$ *and with a* 3^{rd}*-degree polynomial for* $3.5 < z_{\mathrm{Ly}\alpha} < 4.7$.

several pixels, would generate additional correlations. We evaluate the impact of pixel masking by applying the same procedure on mock spectra that include neither sky lines nor DLAs, and we measure the ratio :

$$b(k,z) = \frac{P_m(k,z)}{P_u(k,z)}, \tag{4.11}$$

with P_m the masked power spectrum and P_u the unmasked power spectrum. In order to only evaluate the masking impact, both power spectra are computed using the generated values for the quasar continuum $C_q(\lambda)$ and for the mean transmitted flux $\overline{F}(z_{\mathrm{Ly}\alpha})$.

The impact of the sky emission line masking is illustrated in Fig. 4.9. No strong sky line enters the forest for $2.7 < z_{\mathrm{Ly}\alpha} < 3.3$, which explains why no bias is observed in the corresponding redshift bins. The largest bias occurs for the $z_{\mathrm{Ly}\alpha} = 4.2$ redshift bin, since it is the one with the largest number of sky lines in the forest. For most of the impacted bins, we observe an underestimation on large scales and an overestimation on small scales. At first glance, the result is surprising, as we would naively expect masking to yield a loss of power. However, we can model the effect of pixel masking as the convolution of the unmasked power spectrum by the squared Fourier Transform of the masking function. The masking function being either zero or one, according to whether the pixel is masked or not, it can be expressed as a sum of rectangular functions. As our initial power spectrum is decreasing with k, it appears natural to observe an excess of power on large k-modes (i.e., small-scales).

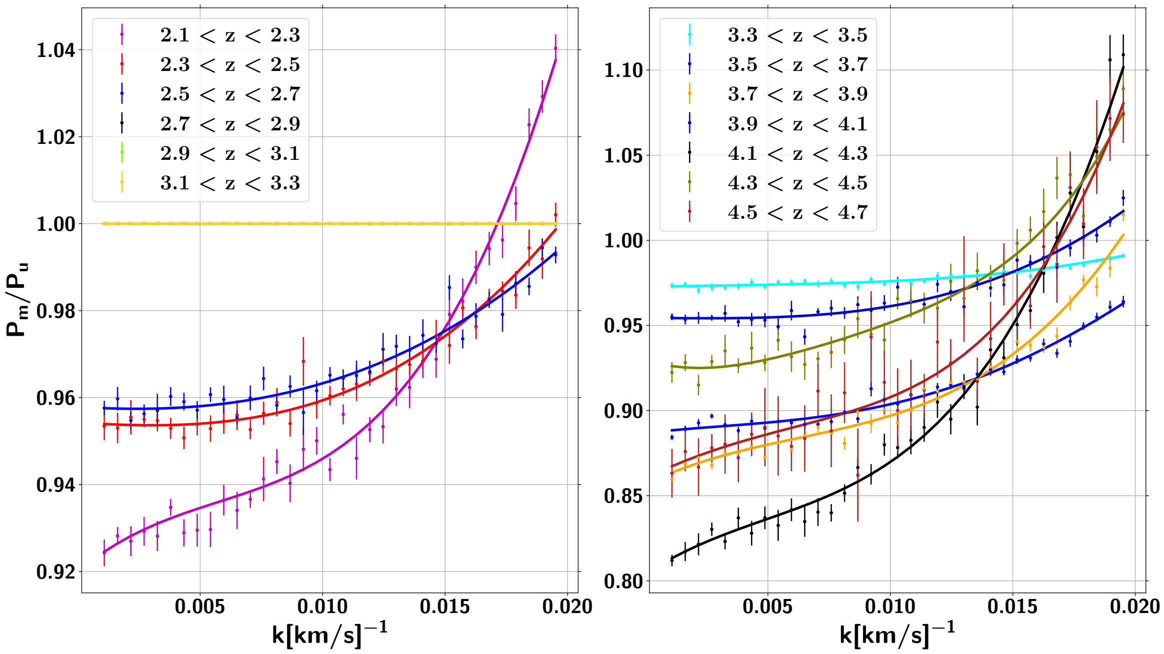

FIGURE 4.9 – *Ratio of the masked power spectrum P_m to the unmasked power spectrum P_u, due to sky emission line masking, for all thirteen redshift bins from $z_{\text{Ly}\alpha} = 2.2$ to 4.6. No strong sky line enters the forest in $2.7 < z_{\text{Ly}\alpha} < 3.3$, resulting in no correction in this redshift range. The bias in $4.3 < z_{\text{Ly}\alpha} < 4.7$ and $3.5 < z_{\text{Ly}\alpha} < 3.7$ is modeled by a 4^{th}-degree polynomial, and in the other redshift bins by a 3^{rd}-degree polynomial.*

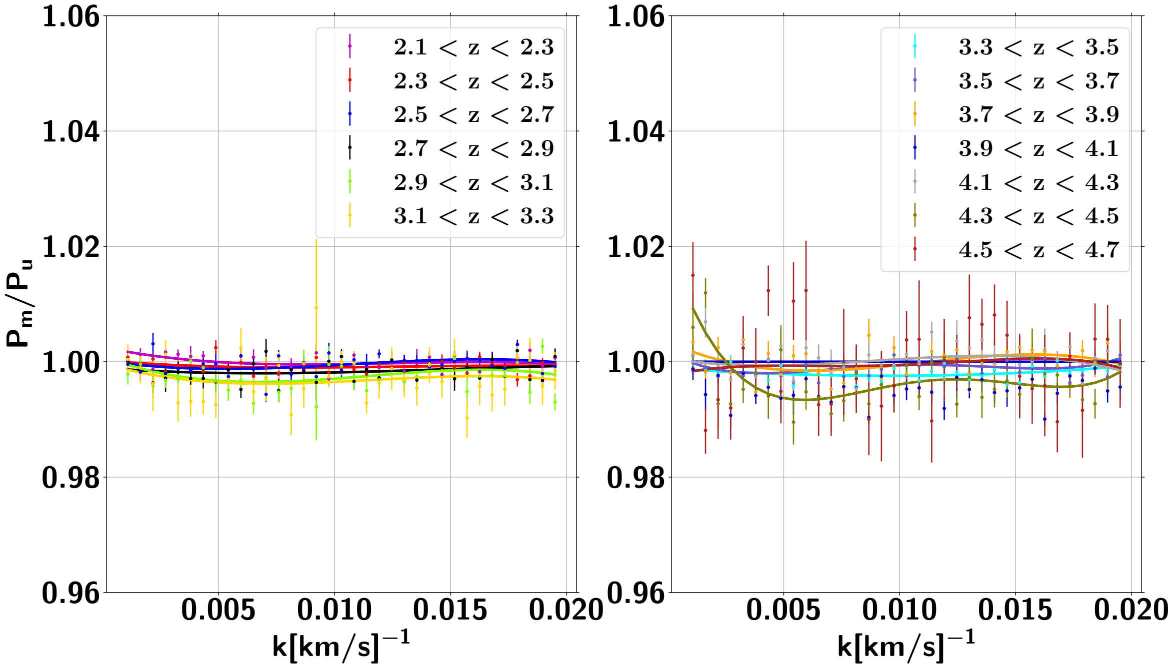

FIGURE 4.10 – *Ratio of the masked power spectrum P_m to the unmasked power spectrum P_u, due to the masking of DLAs, for all thirteen redshift bins from $z_{\text{Ly}\alpha} = 2.2$ to 4.6. The bias in $2.1 < z_{\text{Ly}\alpha} < 4.3$ and $4.3 < z_{\text{Ly}\alpha} < 4.7$ is modeled by a 3^{rd} and a 4^{th}-degree polynomial, respectively.*

The impact of the DLA masking is presented in Fig. 4.10. As DLAs are arbitrarily distributed in wavelength and in strength, and furthermore impact only a fraction of the forests at a given redshift, their masking induces a significantly smaller effect than that of sky lines. The low statistics of DLAs is responsible for the scatter, in particular at high redshift. The measured bias varies with redshift along with the fraction of forests, in our selection, affected by DLAs. The fraction is less than 1% at low redshift, around 5% at intermediate redshifts and of 15% at $z = 4.6$. The redshift evolution is in agreement with measurements in hydrodynamical simulations by Rogers et al. (2018) for instance, where their Tab. 1 shows an increasingly large fraction of their Lyα spectra contaminated by DLAs as the redshift increases. This result is true for all categories, from small Lyman limit systems to strong DLA absorbers. The trend is explained by the increasing total cross-section of DLAs with redshift.

Tab. 4.3 summarizes the sources of bias identified in the analysis. The final power spectra are corrected by the corresponding k- and z-dependent correction functions. The related systematic uncertainties associated to each of these corrections are discussed in the next section.

QSO continuum	1.02 to 1.05
Masking of sky lines	0.82 to 1.10
Masking of DLAs	0.99 to 1.00

Table 4.3 – *Maximum range of the corrections introduced at different steps of the analysis*

4.4 Systematics

As we explained in the previous two sections, going from equation (4.7) (how P_{raw} is derived from observational quantities) to the final measurement of P_{1D} requires selections and power spectrum corrections at several stages of the analysis. These corrections and the impact of the selections are each determined with their own degree of precision, from which we infer a k- and z-dependent systematic uncertainty on the measurement of P_{1D}. We identify eight systematic uncertainties :

- Measurement of the quasar spectrum continuum
- Measurement of the quasar spectrum noise level
- Measurement of the spectrograph spectral resolution
- Measurement of the power spectrum in side bands
- Effect of masking of the sky emission lines
- Effect of masking of the DLA absorbers
- Effect of the completeness of the DLA catalog
- Effect of the completeness of the BAL catalog

I now briefly describe each of these systematic uncertainties. Their impact is summarized in Fig. 4.11 and Fig.4.12. The systematic uncertainties we have identified are expected to be uncorrelated and can therefore be added in quadrature.

As explained in Sec. 4.3.2, the correction related to the determination of the continuum is validated by assessing that consistent power spectra are obtained after application of the correction for the continuum estimate, whichever form is used for the quasar-dependent term of the continuum function. The agreement is at the 1% level. We assign a 30% uncertainty

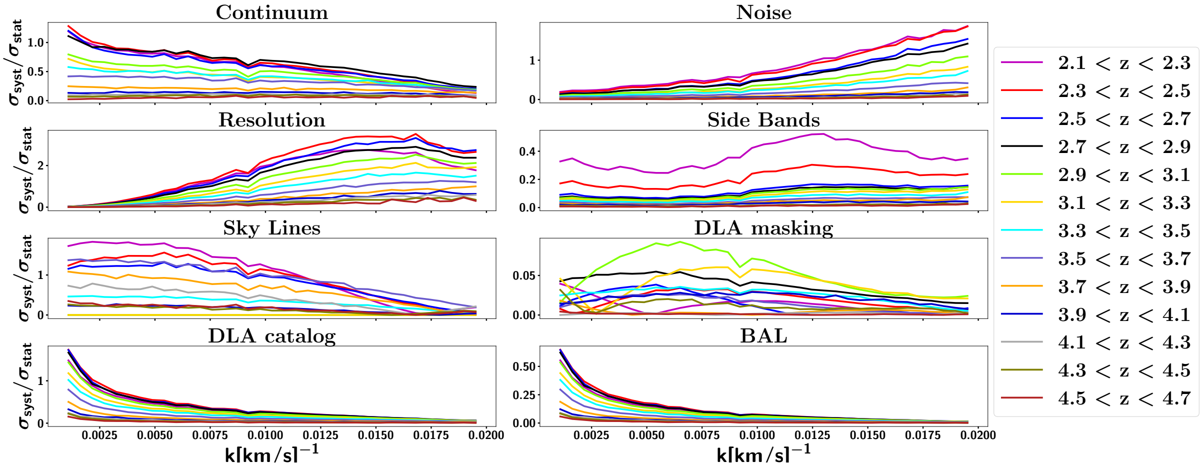

FIGURE 4.11 – *Ratios of the systematic to the statistic uncertainties, as a function of redshift and wave number, for the eight identified sources. From left to right and top to bottom are illustrated the uncertainty ratios from the continuum estimation, the noise level, the spectral resolution, the side bands power spectra, the sky lines masking, the DLA masking, the DLA residual effects and the BAL features.*

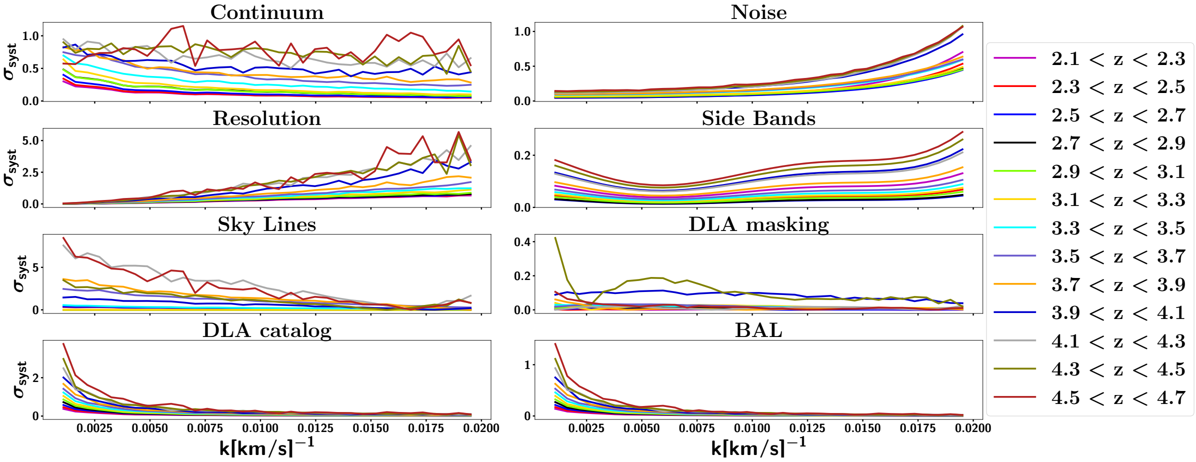

FIGURE 4.12 – *Absolute values of the systematic uncertainties, in* $\mathrm{km \cdot s^{-1}}$*, as a function of redshift and wave number, for the eight identified sources, in the same format as Fig. 4.11.*

on the correction measured, which leads to an uncertainty of comparable magnitude to the aforementioned agreement. This 30% comes from the following assumption. As a conservative choice on unknowns that could affect the value of the correction, we consider a shift randomly ranging between 'no correction' and '100% of the correction', which we describe by a uniform distribution between 0 and 1. The standard deviation of the distribution, equal to $1/\sqrt{12} \sim 0.30$, quantifies the spread among the possible values, leading to a systematic uncertainty equal to 30% of the correction.

The quasar spectrum noise level is determined through the procedure described in Sec. 4.2.3. A ratio β different from 1 is an indication of a small discrepancy between the measured noise power spectrum and the one present in the coadded raw spectrum that the SDSS pipeline delivers. We assign a systematic uncertainty on the resulting noise power spectrum equal to the 30% of largest (1- β) term, all redshifts considered. The maximum value is obtained for $z_{\mathrm{Ly}\alpha} = 2.2$, where the noise dominates P_{raw}. Since the noise has a white power spectrum, the spectrograph window function (see equations 4.7 and 4.8) makes the impact of this systematic most significant on large k's.

One of the main systematic uncertainties of this analysis is the knowledge of the spectral resolution R which enters the window function term, $W^2(k, R, \Delta v)$. The spectral resolution is measured in SDSS using arc lamps. From several studies (see Smee et al. (2013); Palanque-Delabrouille et al. (2013)), we derive that the accuracy $\Delta R/R$ on the measurement of R is of order 2%. To take this uncertainty into account, we compute the average resolution $\langle R \rangle$ over the list of quasars that contribute to each redshift bin. The systematic uncertainty on $P_{\mathrm{1D}}(k)$ is then given by $P_{\mathrm{1D}}(k) \cdot (2k^2 R \Delta R)$. The quadratic k-term makes the large k-modes more affected by this uncertainty. The mean resolution $\langle R \rangle$ also varies with redshift, from $81\,\mathrm{s}\cdot\mathrm{km}^{-1}$ to $65\,\mathrm{s}\cdot\mathrm{km}^{-1}$ with larger values for lower redshifts. This redshift-dependence induces a larger impact for low-z bins.

The power spectrum in the side bands, defined in Sec. 4.2.4, is used to estimate the power spectrum of the metal absorption and to correct for residual effects of the pipeline. However, the accuracy of these corrections are limited by the numbers of quasars with side bands in the relevant observed wavelength range. Therefore we propagate, as systematic uncertainties, the statistical errors on the determination of $P_{\mathrm{SB}_1}(k)$. As shown in Fig. 4.7, the shape of the power spectrum is obtained, for each redshift bin, from the product of a universal sixth-degree polynomial derived from the average shape for all the quasars, and a first-degree polynomial in which both parameters are free. We vary the shapes according to the statistical errors on the latter parameters to estimate the systematic uncertainties. The largest systematic uncertainties are obtained at high k for the three redshift bins that contain the overlap between the two arms of the spectrograph : $5800 < \lambda < 6350\,\mathrm{\AA}$.

The level of uncertainty on the correction of sky line masking, computed by means of mock spectra, varies with the amplitude and shape of the input power spectrum $P_g(k)$ and with the value of the spectral resolution. These two input parameters were therefore chosen, in the mocks, to reproduce as well as possible the measured values of these inputs at all redshifts. The spectral resolution and input power spectrum were varied within observational limits ; the measured variations of the derived corrections were at the level of 3% and 5%, respectively. To include both dependences, we assign a conservative overall 30% uncertainty to the correction for sky line masking. As shown in Sec. 4.3.3, the largest effect occurs where the bias is largest, e.g., at redshift $z_{\mathrm{Ly}\alpha} = 4.2$ for low k's. In contrast, there are no sky line systematics for redshifts $2.7 < z_{\mathrm{Ly}\alpha} < 3.3$, since such forests contain no strong sky lines.

Like done in the masking of the sky lines, we assign a 30% systematic uncertainty associated to the correction for the masking of DLA absorbers. Because the DLA masking yields at most a 1% correction (cf. Fig. 4.10), the related systematic is sub-dominant compared to all others. The DLA correction shows almost no dependence on k, and the redshift dependence is explained by the increasing percentage of contaminated forest.

The residual effect of unmasked DLAs was not taken into account in the systematics budget of PYB13 analysis. Here, we use the automated DLA catalog of N12 Noterdaeme et al. (2012), as was done in Bautista et al. (2017); du Mas des Bourboux et al. (2017). We compute a systematic uncertainty associated to this sample from the data themselves, using an alternative DLA catalog from P18 (Parks et al., 2018). Fig. 4.13 displays the distribution of the column density $\mathrm{N_{HI}}$ for both catalogs. Both catalogs were optimized to identify DLAs, i.e., absorbers with $\mathrm{N_{HI}} \geq 10^{20.3}\,\mathrm{cm}^{-2}$. P18 has a lower minimum column density and includes many more sub-DLAs and weak DLAs, which still have a significant impact on the power spectra as demonstrated in Rogers et al. (2018). P18 contains 4,419 DLAs in the selected forest sample and N12 contains 2,105 DLAs. We compare the resulting power spectrum to that obtained when masking with the superset of P18 *and* N12. When possible, we select the $\mathrm{N_{HI}}$ from N12 to provide as consistent a comparison as possible. As shown in Rogers et al. (2018), the impact of DLAs on the power spectrum strongly depends on the $\mathrm{N_{HI}}$ of the absorbers considered. Fig. 4.14 illustrates the effect of the additional DLAs of P18, integrated over all forest redshifts. Our results indicate a rise of a few percent on the largest scales ($k < 0.003\ \mathrm{s \cdot km^{-1}}$), in qualitative agreement with Rogers et al. (2018). We assign an uncertainty equal to 30% of the ratio ($P_{N12}(k)/P_{P18 \cup N12}(k)$-1).

We reject from the analysis all quasars exhibiting BAL features, identified by a non-zero BI_CIV flag. We consider that the automated procedure identifies the 80% largest BALs with high efficiency, but could be incomplete for the 20% faintest ones, i.e., those with BI_CIV $<$ $170\,\mathrm{km \cdot s^{-1}}$. We thus compute the ratio $P_{\mathrm{BI\_CIV}>0}(k)/P_{\mathrm{BI\_CIV}>170}(k)$, and we assign a systematic uncertainty equal to 100% of this ratio. Unidentified BALs mostly affect the power spectrum on large scales, but the effect remains sub-dominant.

4.5 Results

Using the procedure described in the previous sections, we compute the 1D power spectrum over 13 redshift bins from $z_{\mathrm{Ly}\alpha} = 2.2$ to 4.6, and over 35 modes from $k = 10^{-3}$ to $k = 0.02\ \mathrm{s \cdot km^{-1}}$. The resulting power spectra are presented in Fig. 4.15. The results are in excellent agreement with the one published in PYB13, with no significant shift on any of the points, as also visible in the same figure. The statistical errors are estimated in two ways, either from the rms of the distribution of the values of the power spectrum for a given k and $z_{\mathrm{Ly}\alpha}$ over all contributing forests, or using a bootstrap approach. Both methods yield similar results.

The statistical uncertainties σ_{stat} are reduced by about a factor of two relative to PYB13 at all redshifts, due to the approximately four-fold increase of the selected quasar sample. Fig. 4.16 shows this extreme decrease of the statistical uncertainties, with the statistical errors at all redshifts for PYB13 and for this analysis, rebinned in two wave numbers with one for large scales, i.e $k < 0.01\,\mathrm{s \cdot km^{-1}}$, and one for small scales, i.e. $k > 0.01\,\mathrm{s \cdot km^{-1}}$. The systematic uncertainties σ_{syst}, in contrast, are increased by a about a factor of two, due to a more thorough investigation

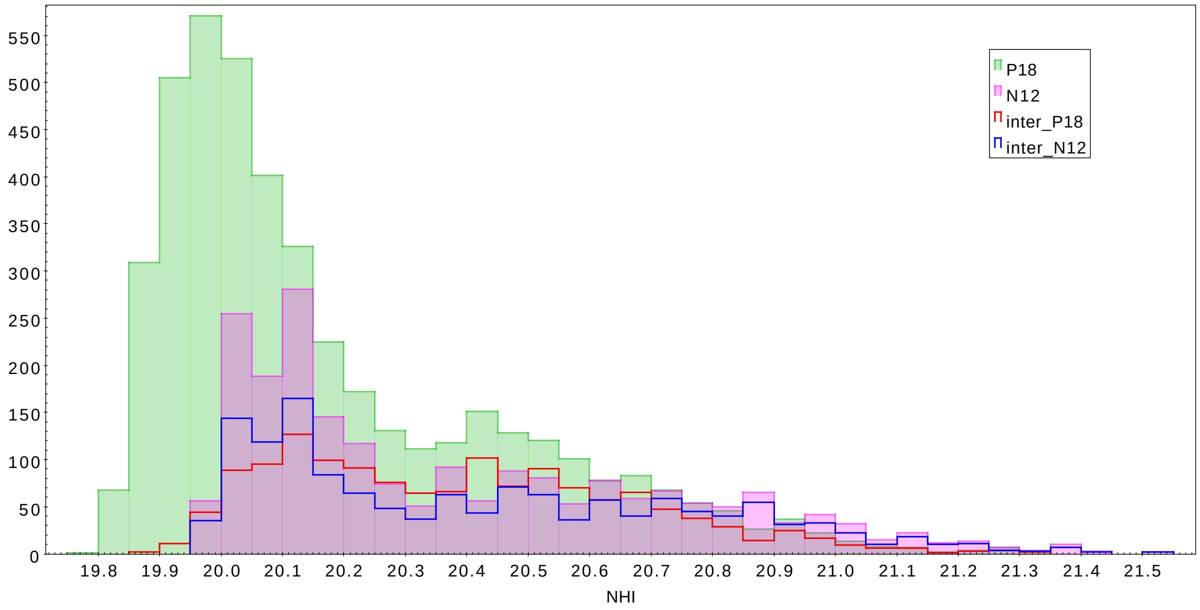

Figure 4.13 – *Distribution of* $\log \mathrm{N_{HI}}/(\mathrm{cm}^{-2})$ *for N12 in magenta and P18 in green. The solid lines show the* $\mathrm{N_{HI}}$ *distribution of DLAs that are present in both catalogs. This illustrates the difference in the* $\mathrm{N_{HI}}$ *estimation for common objects.*

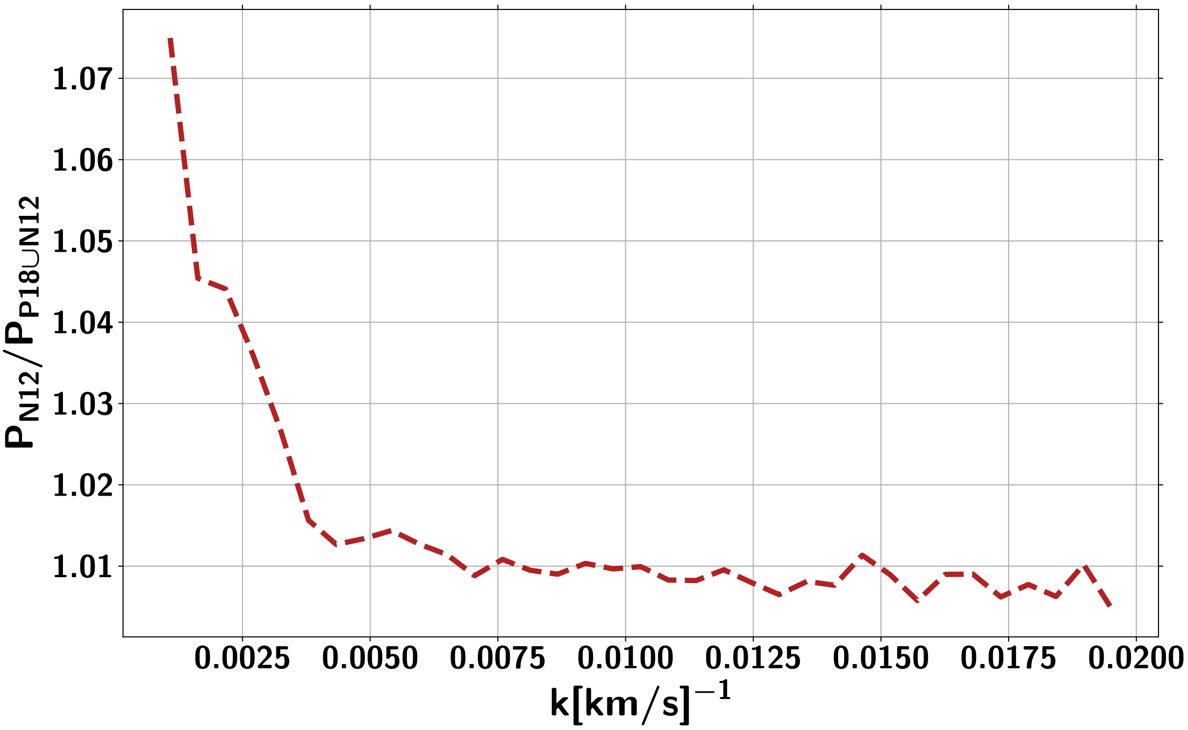

Figure 4.14 – *Ratio of the power spectra using N12 to the union of P18 and N12 as DLA catalogs. From the distribution shown in Fig. 4.13 it can be interpreted as the ratio of sub- and small- DLAs power spectra to the uncontamined power spectra.*

of the possible sources of systematics that affect the measured power spectrum. The major difference arises from the study of the impact of the possible incompleteness of the BAL and DLA catalogs. Although the contribution of the resulting uncertainty is larger for large redshifts where P_{1D} is larger (cf. Fig. 4.12), the remarkably small value of the statistical uncertainty at small redshift makes the relative contribution of $\sigma_{\rm syst}$ more important at low redshift (cf. Fig. 4.11), when all resdshifts in PYB13 were limited by statistical size of the quasar sample. We also measure a k-dependent bias, and hence a k-dependent systematic uncertainty, resulting from the procedure used to determine the quasar continuum. This feature was not observed in the previous method, which, however, was less sophisticated and did not include a quasar-dependent term. The systematic uncertainty related to the correction for the noise power is slightly reduced compared to PYB13. Other error contributions, such as systematics related to side bands, sky line and DLA masking, are similar to what was measured before.

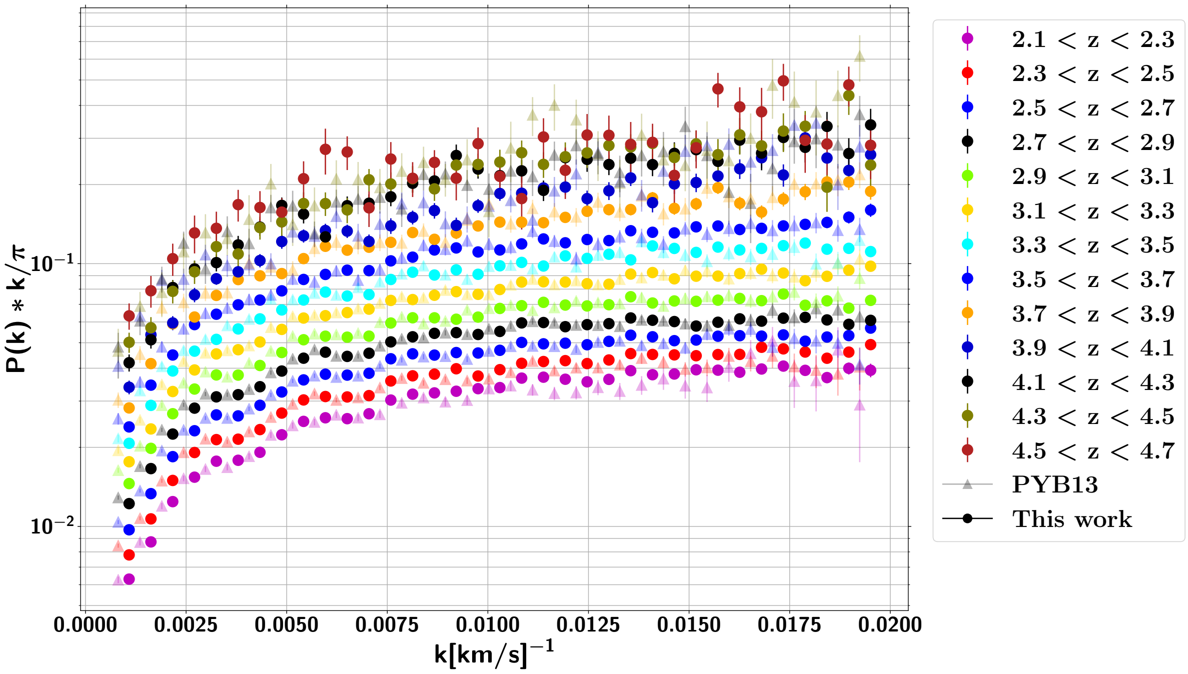

FIGURE 4.15 – *The 1D Lyα forest power spectrum. The measurements from PYB13 are shown in light colors, slightly shifted to smaller k's for the better clarity. Error bars are statistical only.*

The agreement between the two analyses can be examined with a pull distribution, i.e. the distribution of the difference in $P(k, z)$ divided by the statistical error on that difference. Because a fraction of the quasars selected in the DR9 analysis of PYB13 are included in the present work, the combined statistical error overestimates the error on the difference. However, given the increased size of the data set and more optimized selection of the present analysis (leading to a factor of two reduction of $\sigma_{\rm stat}$), the overestimate is 10% at most. The pull distribution has a mean of 0.04 ± 0.05, thus indicating excellent agreement on average, and a standard deviation of 1.18 ± 0.10. The spread slightly exceeds the one expected from purely statistical effects. This result is to be expected, since some steps of the analysis procedure are affected by different systematic uncertainties. To obtain a qualitative insight of the impact of these systematics, a reasonable compromise is to add in quadrature the systematic uncertainties from the present analysis only : we include sources of biases between the two pipelines while not double-counting

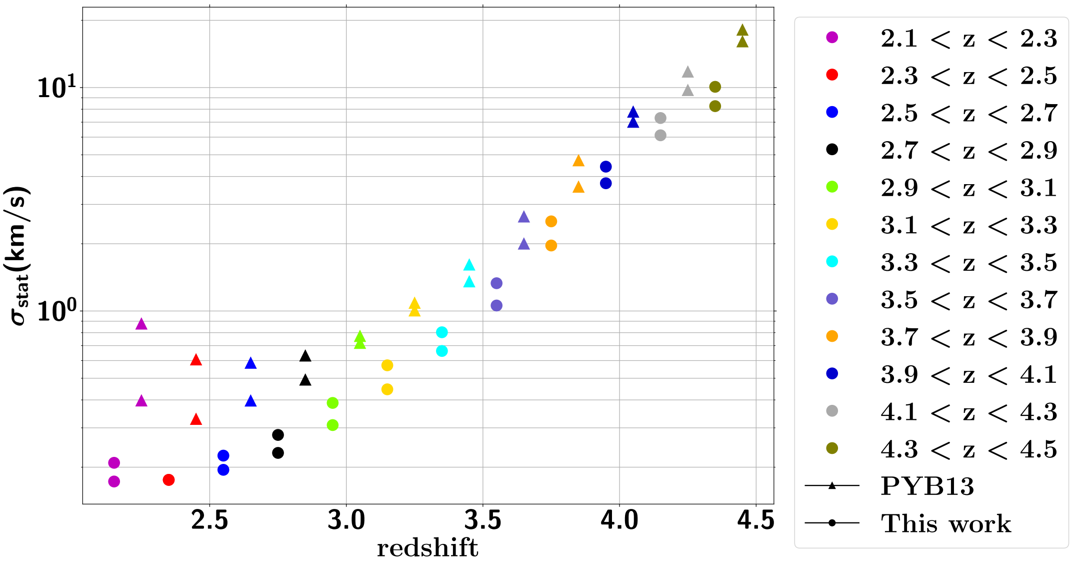

FIGURE 4.16 – *Statistical errors for all common redshifts between this work (circles) and PYB13 (triangles), rebinned in two wave numbers with one for large scales, i.e $k < 0.01\,\mathrm{s\cdot km^{-1}}$, and one for small scales, i.e. $k > 0.01\,\mathrm{s\cdot km^{-1}}$. Statistical uncertainties are larger at high k's, therefore the top points for each redshifts and for the two analysis is always the statistical error at small scales.*

the contributions that are common to both. This approach reduces the standard deviation of the ratio to 1.00, indicating that the differences between the power spectra resulting from PYB13 and the present analysis are well explained by their uncertainties.

Finally, Fig. 4.17 and Fig. 4.18 display the correlation matrices measured for each of the 13 redshift bins, smoothed by second-order polynomials both along and across lines parallel to the diagonal elements of the matrix. The rms and the bootstrap approaches yield similar results. The correlation coefficients are 15 to 20% at most at low redshift ($z_{\mathrm{Ly}\alpha} < 3.0$) and on large scales ($k < 0.01\,\mathrm{s\cdot km^{-1}}$), and quickly decrease to values below 5% otherwise. There is negligible correlation between redshift bins because the forest range (cf. Sec. 4.1.2 for the wavelength coverage) has a redshift extension $\Delta z = 0.2$ at most. Each forest thus contributes to a single redshift bin. Moreover, the three sub-forests of a given quasar are processed independently to avoid induced correlations between them. We check this assumption with a bootstrap approach. Fig. 4.19 shows the result for the $z = 2.2$, 2.4, and 2.6 redshift bins that present the strongest correlation on large scales. As expected, distinct redshift bins do not exhibit any correlation.

Tab. 4.4 provides an extract of the measured power spectrum, as well as statistical and systematic uncertainties. The full table and correlation matrices are available online attached to the publication in Chabanier et al. (2019), so it can be used by other goups for different cosmological or astrophysical purposes.

4.6 Summary and prospects

I this Chapter I presented the measurement of the 1D power spectrum of the Lyα forest that I led during the first part of my PhD using high-quality 43,751 quasar spectra covering a large redshift range from $z_{\mathrm{Ly}\alpha} = 2.2$ to 4.6.

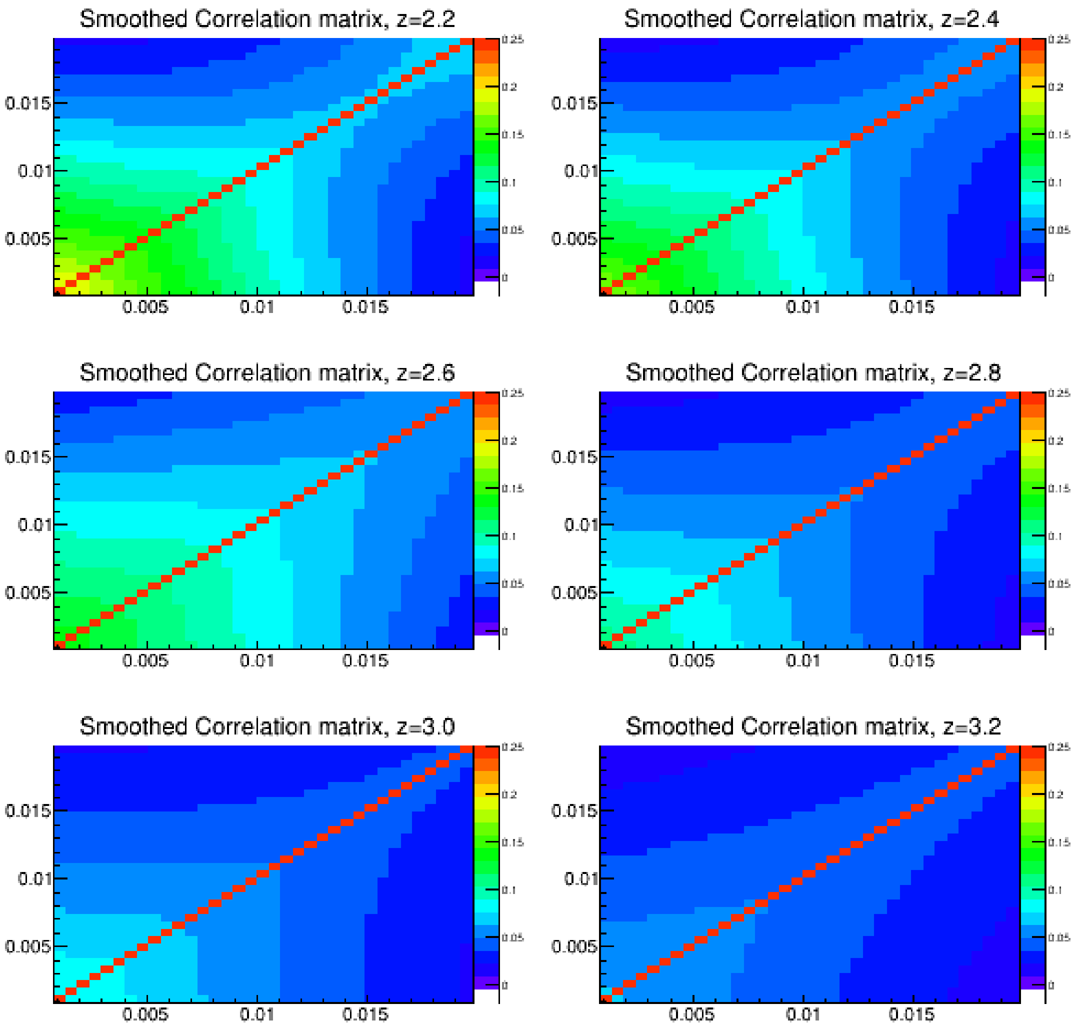

FIGURE 4.17 – *Correlation matrices between k-modes for the redshift bins from $z = 2.2$ to $z = 3.2$. Axes are k modes in* $\mathrm{s \cdot km^{-1}}$. *The color range is chosen to saturate at a correlation of 25%.*

$z_{\mathrm{Ly}\alpha}$	k	$P_{Ly\alpha}$	σ_{stat}	P_{noise}	P_{SB}	$\sigma_{\mathrm{sys}\,1}$	$\sigma_{\mathrm{sys}\,2}$	$\sigma_{\mathrm{sys}\,3}$	$\sigma_{\mathrm{sys}\,4}$	$\sigma_{\mathrm{sys}\,5}$	$\sigma_{\mathrm{sys}\,6}$	$\sigma_{\mathrm{sys}\,7}$	$\sigma_{\mathrm{sys}\,8}$
2.2	0.00108	19.2561	0.2527	2.5551	3.0573	0.3008	0.0491	0.0060	0.0828	0.4361	0.0100	0.3946	0.1474
2.2	0.00163	17.4875	0.2152	2.5764	2.5723	0.2238	0.0495	0.0122	0.0752	0.3859	0.0075	0.2441	0.0893
...						...							

TABLE 4.4 – *Measured power spectrum $P_{Ly\alpha}$ in km$\cdot$s^{-1} for each redshift bin $z_{\mathrm{Ly}\alpha}$ and scale k in* $\mathrm{s \cdot km^{-1}}$. *Also listed are the statistical uncertainty σ_{stat}, the noise power P_{noise}, the side-band power P_{SB}, and each of the systematic uncertainties from the estimate of (1) continuum, (2) noise power, (3) spectrograph resolution, (4) side band power, (5) sky line masking, (6) DLA masking, (7) DLA catalog completeness, (8) BAL catalog completeness. Uncertainties, noise and side-band power are in* $\mathrm{km \cdot s^{-1}}$.

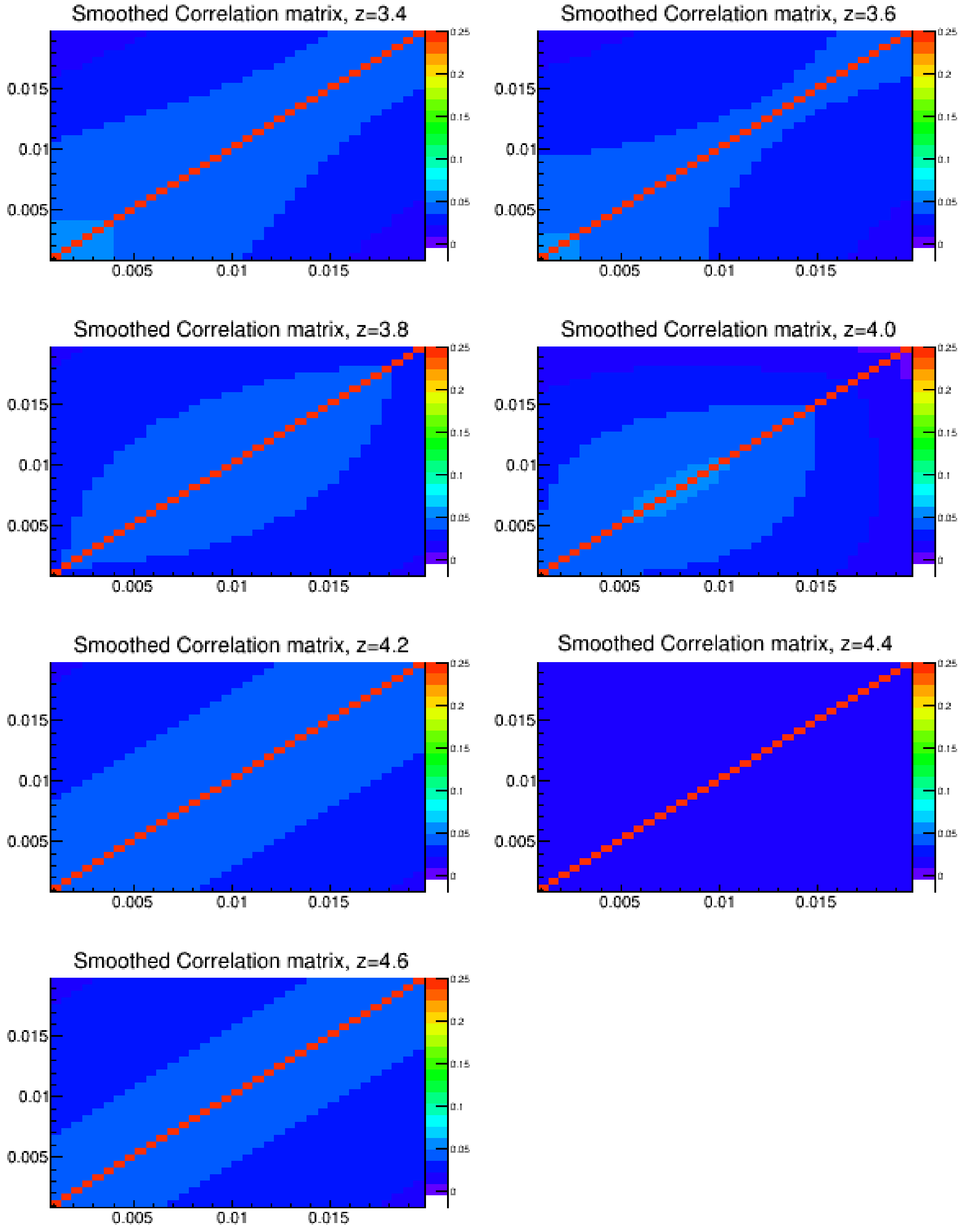

Figure 4.18 – *Correlation matrices between k-modes for the redshift bins from $z = 3.4$ to $z = 4.6$. Axes are k modes in $\mathrm{s \cdot km^{-1}}$. The color range is chosen to saturate at a correlation of 25%.*

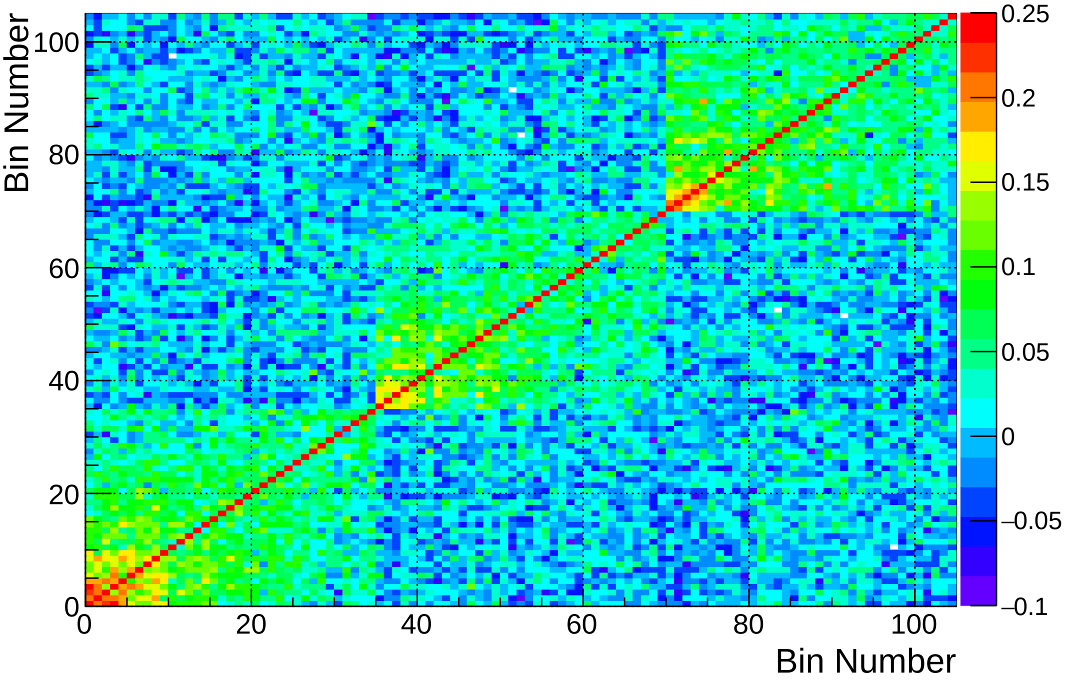

FIGURE 4.19 – *Correlation matrix between* k*-modes for the* $z = 2.2$, 2.4 *and* 2.6 *redshift bins. For each redshift, the power spectrum is measured over 35* k *bins. The axes of this correlation matrix therefore have* 3×35 k *bins. The diagonal* 35×35 *sub-matrices correspond to the first three matrices of Fig. 4.17. The color range is chosen to saturate at a correlation of 25%.*

The increased data set allows the addition of a new redshift bin at $z_{\mathrm{Ly}\alpha} = 4.6$ compared to PYB13. Despite the lower statistical precision of this bin, it carries useful information from an earlier epoch in the history of the Universe, where the clustering is less affected by non linearities than at lower redshift. This redshift bin is therefore highly valuable to constrain dark matter properties.

This unprecedented statistical power of the data requires to perform a careful investigation of all the systematic errors and their sources. We identify eight sources of systematics related to the analysis procedure that we finely studied using mocks tuned to match the level of the data power spectrum. The major sources of uncertainty on small scales come from the precision on the determination of the spectrograph resolution, and from the estimation of the noise power. The latter largely dominates over the cosmological power in particular at low redshift. These two issues should be improved with the next generation WEAVE-QSO (Pieri et al., 2016) and Dark Energy Spectroscopic Instrument (DESI) (Collaboration et al., 2016) projects, which will have almost twice the spectral resolution and a higher signal-to-noise ratio for $z_{\mathrm{qso}} > 2.1$ quasars. On large scales, the dominant source of uncertainty arises from the incompleteness of the BAL and DLA catalogs. The data quality in next generation surveys will improve their identification. A part of my thesis was also dedicated to use automated algorithm to detect DLAs in the sixteenth data release (DR16) of the SDSS (Ahumada et al., 2019) and generate a public catalogue (Chabanier et al. in prep) which is useful not only for $P_{\mathrm{Ly}\alpha}$ studies but also for Lyα auto-correlation or cross-correlation between DLAs and quasars or DLAs and Lyα. With a three to four-fold increase in quasar number density with the future spectroscopic surveys, it will be possible to further tighten the selection of the quasar spectra and reduce the contamination from systematic biases.

This measurement achieves an unprecedented level of precision, it even reaches the percent-level accuracy at low redshifts. The inference of cosmological parameters at this level of precision needs to rely on state-of-the-art hydrodynamical simulations because the fluctuations probed are in the (midly) non-linear regime, and the imprint of baryonic processes need to be quantified and disentangled from that of cosmological parameters. Precision cosmology with the Lyα forest can shed light on the nature of dark matter particles, neutrinos or the accelerating expansion of the universe, but only if dedicated simulations reach the same percent-level accuracy. In the next Chapters, I will present the use of numerical simulations in cosmology, and in particular to precisely model the Lyα forest in order to have the adequate accuracy in our theoretical predictions to make robust scientific interpretations.

5

Numerical simulations in Cosmology

Contents

In order to constrain cosmological models from observational data, using e.g. the two-point correlation function of galaxies or the power spectrum of the Lyα forest, these have to be confronted with theoretical predictions exploring the cosmological parameter space. From statistical comparisons, we can estimate the best fitting cosmological parameters and the associated confidence intervals. This Chapter aims at giving the basic tools to precisely model the Lyα forest by means of cosmological hydrodynamical simulations.

Standard approaches often use analytical recipes based on first principles computing the evolution of perturbations. However such computations become increasingly difficult to perform in the non-linear regime. Non-linear corrections can be included using standard perturbation theory (SPT), regularization perturbation theory (RPT) or effective field theory. But a trade-off between the accuracy of predictions and the precision on inferred parameters has to be made. Also, a cutoff scale has to be chosen below which the approximations break down preventing to make any benefits from the otherwise so valuable small-scale data. Non linearities is a well-posed numerical problem that that can be solved at all scales with high accuracy using numerical simulations. Because the Lyα forest probes scales at the transition from the linear to the non-linear regime ($k \sim 1\ \mathrm{Mpc} \cdot \mathrm{h}^{-1}$), and also because we observe light flux from quasar we cannot estimate theoretical predictions using analytical recipes. Hence, we have to rely on numerical simulations.

Numerical simulations are ab-inito approaches, going from initial density fluctuations, which simulate the evolution of structures and their different components, dark matter, neutrinos and the ordinary baryonic matter, given a cosmological model, a theory of gravity and relevant physical processes. They afford to efficiently solve gravitational instabilities and to include a realistic treatement of complex physical processes related to the baryonic component in order to precisely probe the impact of massive neutrinos and warm dark matter.

The intergalactic gas giving rise to the Lyα forest is the result of the complex interplay between large-scale structure evolution driven by the dark-matter density fluctuations, and small-scale baryonic physics. Indeed, astrophysical processes such as star formation, stellar and AGN feedback injecting energy in the ambient medium strongly impact the thermal state and distribution of gas in the IGM. As such, the modeling of the Lyα forest is intrinsically a multi-physics and multi-scale problem, which requires a huge dynamical range, between the kpc scale and few hundreds of Mpc, i.e. 10^5. Not only does it require to accurately model the IGM physics, but it also demands to have detailed galaxy-formation and evolution models. If dark matter only simulations, solving only gravitational instabilities, is a well-posed numerical problem, the details of the formation and evolution of galaxies are much less well understood. The main reason of the difficulty in modeling the baryons is due the complexity and non-linearity of underlying physical processes. The collisional nature of baryonic gas makes it much more complicated to model than collisionless matter.

Sec. 5.1 presents the general numerical methods to simulate the universe, going through the cosmological framework in Sec. 5.1.1, initial conditions in Sec. 5.1.2, techniques to solve gravitational dynamics for cold dark matter in Sec. 5.1.3 and massive neutrinos in Sec. 5.1.4, hydrodynamics of collisional gas in Sec. 5.1.5 and extra baryonic physics needed to properly model the galaxy evolution and consequently the Lyα forest in 5.1.6. A special emphasis is put on supernovae and AGN feedbacks as I studied these two features in details (see Chapter 6) Finally, Sec. 5.2 describes the Adaptative Mesh Refinement hydrodynamical code RAMSES that I extensively used in my thesis.

5.1 General numerical methods

Simulating the universe requires to solve the equations of the main matter components, cold dark matter, massive neutrinos and baryonic gas, which are discretized and evolved forward in time. This is done within a cosmological framework including physical models for gravity and dark energy, responsible for the accelerated expansion. The dark matter component, forming the backbone of the cosmic web and massive neutrinos follow the equations of collisionless gravitational dynamics. The gas component is described through the set of the Euler equations for adiabatic evolution. Finally, various astrophysical processes are included in order to simulate realistic galaxies, often with effective sub-resolution prescriptions.

5.1.1 The cosmological framework

Cosmological simulations are often performed in boxes with periodic boundary conditions to mimic the large-scale homogeneity and isotropy so that the cosmological principle is well accounted for. They aim at simulating large volumes with box sizes from a few $(\mathrm{Mpc}/h)^3$ to a few $(\mathrm{Gpc}/h)^3$. The smallest resolution element is usually much poorer than in galactic-scale simulations, but these simulations can statistically study the properties of large-scale structures as a function of redshifts. Typically, cosmological simulations can go down to the kpc scale, while galactic-scale simulations can easily go below the pc scale. A cosmological model has to be chosen via a set of adequate cosmological parameters. Various observations, ranging from the CMB, to galaxy clustering, weak lensing or the Lyα forest, can all be explained with only the minimal six-parameters ΛCDM model. About 95% of the energy budget is composed of dark energy and dark matter while baryons make up for the 5% remaining. In order to infer cosmological parameters via statistical inference techniques with observational data, we have to construct a grid of simulations efficiently covering the cosmological parameter space.

The cold dark matter (CDM) model assumes that dark matter is cold, with negligible random motions when decoupled from other matter, and collisionless. This standard dark matter model is implemented with Newtonian gravitational dynamics (described in Sec. 5.1.3). However, the CDM model suffers from discrepancies with astrophysical observations at small scales (below the Mpc). These are the so-called *small-scale controversies* presented in Chapter 1. Alternative models for dark matter are implemented to test if they conserve the accurate CDM large-scale predictions and reconcile with observations on small scales. The warm dark matter (WDM) model is studied is this thesis and fully presented in Sec. 1.3.3 in Chapter 1. Simulations implementing WDM models are typically evolved with the same numerical methods than for the CDM model but with modified initial conditions to mimic the suppression of power at small scales induced by the free streaming of the WDM particle (standard set up are described in Sec. 5.1.2). We use different transfer functions for varying the WDM particle masse, where the small-scale modes are suppressed compared to the transfer function of the CDM. Other exotic models of dark matter have also been implemented, e.g. with fuzzy dark matter (Mocz and Succi, 2015; Nori and Baldi, 2018; Zhang et al., 2018) or self-interacting dark matter (Vogelsberger et al., 2012; Peter et al., 2013; Fry et al., 2015; Elbert et al., 2015), but are not studied in this thesis.

The acceleration of the expansion of the universe is accounted for with the cosmological constant Λ. Alternative models for dark energy can also be implemented to be confronted with observations. For instance, dynamical dark energy is easily implemented by modifying the computation of the Hubble expansion rate and the growth factor for the initial conditions.

5.1.2 Initial Conditions

Once the cosmological model is defined, initial conditions have to be set. It means that we have to specify perturbations imposed on top of a homogeneous expanding background. This step is very important because errors are greatly amplified by the simulation.

To begin, we have to generate the smooth background ; the distribution of particles that represents the unperturbed system. Standard configurations are grid distribution or glass distribution (Baugh et al., 1995; L'Huillier et al., 2014). A regular grid distribution has the major drawback to presents preferred direction along the axis, leading to artefacts, especially at small scales. The glass distribution is made starting from a random distribution. Then particles are displaced using the opposite sign of gravity until they freeze in comoving coordinates. A random distribution produces the adequate white noise of the homogeneous distribution. But it strongly suffers from Poisson noise and its evolution yields to formation of non-linear structures that should not be there for a homogeneous distribution.

The second step is to displace the particles and set their initial velocities so that the resulting perturbations have the adequate power spectrum given that inflation predicts gaussian perturbations that are completely specified by their power spectrum. The initial power spectrum, just after inflation is generally written as

$$P(k) = A\left(\frac{k}{k_0}\right)^{n_s}, \tag{5.1}$$

where k_0 is the pivot mode, n_s the scalar spectral index and A the amplitude.

Simulations usually start around $z \sim 100$, when perturbations are still linear and not just after inflation for obvious computational cost reasons. The initial power spectrum is corrected with the appropriate transfer function $T(k)$, which depends on the cosmological model, and in particular the dark matter model. It is usually computed with numerical codes such as CAMB[1] (Lewis et al., 2000) or CLASS[2] (Lesgourgues, 2011), or mpgrafic[3] (Prunet et al., 2008),

$$P(k) = A\left(\frac{k}{k_0}\right)^{n_s} |T(k,z)|^2 . \tag{5.2}$$

Particle are slightly displaced and their initial velocity is set to get the appropriate power spectrum. This can be done using the Zel'dovich approximation (ZA) (Zel'Dovich, 1970), where the displacements scale linearly with the growth factor ;

$$\mathbf{r}_i = \mathbf{x}_i - D(t)\nabla\Phi(\mathbf{x}_i), \tag{5.3}$$

$$\mathbf{v}_i = -\frac{dD}{dt}(t)\nabla\Phi(\mathbf{x}_i), \tag{5.4}$$

where $\mathbf{r}_i$ is the perturbated position, $\mathbf{x}_i$ the initial position, Φ the gravitational potential, $\mathbf{v}_i$ the velocity and $D(t)$ the growth rate of linear density fluctuations that depends on the cosmology. However, initial conditions generated with ZA exhibit transients, i.e. excitations of nonlinear decaying modes caused by the failure of ZA to conserve momentum, which hinder the true statistics of the density and velocity fields (Crocce et al., 2006). Therefore the simulation has to be started at sufficient high redshifts so that the transients can decay to negligible amplitude (Efstathiou et al., 1985). The ZA corresponds to first order of the Lagrangian perturbation

---

[1] https://camb.info

[2] http://class-code.net/

[3] http://www2.iap.fr/users/pichon/mpgrafic.html

theory. The approximation can be pushed to higher orders so that the transients are both smaller and decay more rapidly (Bertschinger, 2001; Jenkins, 2010; Hahn and Abel, 2011; Garrison et al., 2016), allowing simulations to start later and save computational time. Baryons are assumed to have the same initial positions and velocities than cold dark matter. Their temperature is often roughly initialized to the CMB temperature. Neutrinos can undergo special treatements that will be described in Sec. 5.1.4.

5.1.3 Gravitational dynamics

The physical model determining the gravitational instabilities leading to the formation and evolution of cosmic structures is given by the equations of motion of collisionless particles (dark matter, neutrinos or stars) of mass m, comoving position $\mathbf{x}$ and momentum $\mathbf{p}$. Stars are accurately described as collisionless particles because the number of stars in galaxies is so large that the Hubble time is clearly subdominant compared to the two-body relaxation time. The Liouville theorem states that the flow of a collisionless system in phase space is incompressible leading to the conservation of its the phase-space distribution $f(\mathbf{x}, \mathbf{p}, t)$. Thus, the time evolution of f is obtained by solving the coupled set of equations composed by the collisionless Boltzmann equation (CBE)

$$\frac{\partial f}{\partial t} + \frac{\mathbf{p}}{ma^2}\nabla f - m\nabla\Phi\frac{\partial f}{\partial \mathbf{p}} = 0 \tag{5.5}$$

and the Poisson equation

$$\nabla^2\Phi = 4\pi G a^2 \int f \mathrm{d}\mathbf{p} \tag{5.6}$$

Here, Φ is the gravitational potential and a the scale factor. The Poisson equation results from Newtonian gravity consideration rather than relativistic gravity. This is a fairly good approximation because linear structure growth is the same in the matter dominated regime and because velocities induced by the non-linear evolution of structures are well below the speed of light. The distribution function is normalized to give the mass density when integrated over the entire velocity space

$$\rho(\mathbf{x}, t) = \int f(\mathbf{x}, \mathbf{p}, t) d^3p. \tag{5.7}$$

To solve the CBE, we can use the method of characteristics, which are the trajectories in phase space along which the distribution function is constant. These trajectories happen to be the newtonian trajectories of fluid elements (Saslaw, 1985). They are given by

$$\frac{d\mathbf{p}}{dt} = -m\nabla\Phi, \tag{5.8}$$

and

$$\frac{d\mathbf{x}}{dt} = \frac{\mathbf{p}}{ma^2} \tag{5.9}$$

The high dimensionality (6+1) of the system makes difficult to solve the complete set of characteristics. A standard approach, called the **N-body** problem, is to sample the phase-space density by an ensemble of N phase-space points $\{(\mathbf{x}_i, \mathbf{p}_i)\}, i = 1..N$ and compute their characteristics, i.e. their trajectories in the gravitational potential. This is equivalent to solving the equation of motion of N gravitating bodies in an expanding universe. It is important to keep in mind that these particles do not have physical sense themselves. They are actually super-particles of stars or dark matter. Typically, in modern cosmological simulation, a dark matter particle encloses 10^8 $\mathrm{M}_\odot$. N determines the mass resolution because this approach discretized the mass and the density fields. It should be large enough to reduce the Poisson noise induced by this sampling. Two-body encounters can violate the collisionless feature of the systems so that artificial two-body relaxation is often observed in such simulations. This leads to artificial large-angle scattering of particles caused by close encounters. It is prevented by using gravitational softening that smooths the density field on small scales, but degrades the force resolution in high-density regions.

Other approaches attempt to directly integrate the CBE-Poisson equation in phase space (Yoshikawa et al., 2013; Alard, 2014; Bardos et al., 2016) and discretize the position and velocity fields on the opposite of N-body methods that discretize the density field. This type of methods is not subject to violation of the collisionless feature of the system. However they are dramatically memory consuming and require a very large computational time. They are more adapted to evolve systems with a small dynamical range, such as the very early universe when perturbations are small or to study globular clusters, so that we can keep a fairly good velocity and position resolution in the 6-dimensional space.

In the following I will describe standard N-body numerical approaches because they are more adaptated to the study of the Lyα forest, which requires a very large dynamical range, ranging from hundreds of Mpc to the kpc scale. N-body methods aim at calculating the forces exerted by the ensemble of particles on one particle in order to update its velocity, which in turn is used to update its position and repeat this for every particles.

- **Particle-Particle** (PP) method is the simplest one as it directly solves the integral form of the Poisson equation via direct summation using
$$\Phi(\mathbf{r}) = -G \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'} d\mathbf{r}'. \tag{5.10}$$
At each time step, the forces on each particle are directly computed by summing the contributions of all remaining particles. Then, the equations of motion (equation (5.8) and equation (5.9)) are integrated to obtain the updated velocities and positions of the particles. It is extremely easy to implement but it is also very computationally intensive with complexity $\mathcal{O}\left(N^2\right)$. This prevents the use of this method with a very large number of particles. Typically, it can be used with up to 10^6 particles, when the largest N-body simulations evolve 10^{12} particles.
- **Tree-code** methods are approximations of the direct summation methods, based on the exploitation of a hierarchical multipole expansion. The speed up is obtained by using, for sufficiently distant particles, a single multipole force, instead of computing every single distance reducing the computational cost to $\mathcal{O}\left(N\log(N)\right)$. In practice, the multiple expansion is based on hierarchical grouping by subdividing the simulation volume in a recursive way. There exists a multitude of ways to organize an ensemble of particles in a volume into a tree, such as the Barnes and Hut octree (Barnes and Hut, 1986) or trees based on nearest neighbour pairing (Jernigan and Porter, 1989). Octree are often implemented in cosmological simulation ; each cubic cell is split into up to eight child cells resulting in a tree-like hierarchy of cubic nodes with the root node containing all particles at its bottom. To compute the force exerted by the total mass distributuion on a given particle, close neighboring particles and distant cells are treated differently. The contribution from nearby particles is accounted for individually, while the contribution from distant particles is accounted for collectively using the mass and center of gravity of coarse cells in the tree. As such, the force resolution can be as high as the PP method in very dense regions.

- **Particle-Mesh** (PM) methods aim to solve the differential form of the Poisson equation (equation (5.6)). They convert the particle ditribution into a density field projected on a grid. On a regular cartesian grid, the Poisson equation is efficiently solved in Fourier space using $-k^2\tilde{\Phi}(\mathbf{k}) = 4\Pi G a^2 \tilde{\rho}(\mathbf{k})$, where $\tilde{\Phi}$ and $\tilde{\rho}$ are the Fourier transforms of the density field and the gravitational potential respectively. The use of fast Fourier Transform (FFT) makes the computation very fast. I will present methods to solve this equation on non-uniform grids in Sec. 5.2, dedicated to the RAMSES code that operates on adaptative grids. Then the gravitational potential is differentiated using a finite-difference approximation to compute its gradient and the force at each grid point following equation (5.8). The lowest-order scheme, the two-point-centered, is often implemented in cosmological simulations. Finally, the forces are interpolated to the particle positions to update their velocity and position. For FFT methods, the complexity is significantly reduced to $\mathscr{O}\left(N_{\mathrm{g}}\log(N_{\mathrm{g}})\right)$, where N_{g} is the number of cubic grid. The particle mass decomposition at the grid nodes represents one of the critical steps for the PM method, both in terms of resolution and computational cost. The mass density at the grid point $\mathbf{r}_g$ can be expressed as
$$\rho(r_g) = \frac{m_p}{\Delta^3}\sum_{i=0}^{N-1} W\left(|\mathbf{r}_g - \mathbf{r}_i|\right), \tag{5.11}$$
where m_p is the mass of one particle, Δ is the spacing and W a suitable interpolation function representing the assignment scheme of the discrete mass distribution. The choice of W is directly related to the accuracy of the approximation. The most commonly adopted interpolations are : the nearest-grid-point (NGP) where the mass of each particle is totally assigned to the nearest grid point only yielding discontinuity in the density field, the cloud-in-cell (CIC) where the mass of each particle is assigned to two nearest points in each directions (8 points in 3D) yielding discontinuity in the density gradient, and the triangular-shaped-cell (TSC) where the mass decomposition involves three-nearest points in each directions (27 in 3D) yielding discontinuity in the second derivative of the density field. Higher-order prescriptions decrease the amount of noise but are more computationally expensive. The major flaw of this method is the loss of accuracy and resolution because it does not allow to follow close interactions on small scales (i.e. on subgrid distances).
- **Hybrid methods** combine some of the above methods to exploit either their accuracy or their high computational speed. **P^{3}M** methods combines PP and PM approaches. Large-scale interactions are implemented with the PM method, while PP method is used on small scales to increase the accuracy of the PM method limited by the grid size. The **TreePM** approach implements the PM methods for large-scale interactions and Treecode methods for small-scale ones.

Dark-matter only simulations have played a significant role in cosmology, in particular to understand the formation of the large-scale structures in the universe. In the 1974, the authors in Press and Schechter (1974) used 1,000 particles randomly placed in a sphere, evolved with N-body scheme, to study the formation and evolution of clusters of galaxies. They tested whether the gravitational instability scenario could account for the galaxy clustering that was quantified with different statistical estimators, in particular the two-point correlation function. Ever since, the field of *computational cosmology* has bloomed. The exponential growth of computational power and the developement of new and more efficient algorithms have made it possible to increase both the spatial and mass resolution in simulations. We can for instance note the Eu-

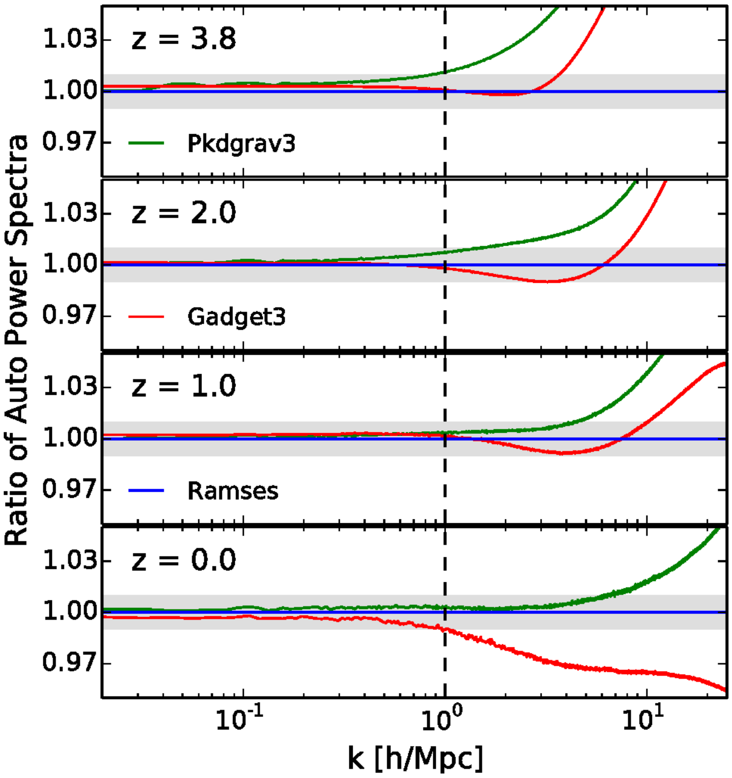

FIGURE 5.1 – *Comparison of auto power spectra from the three different N-body codes at different redshifts. Green lines correspond to Pkdgrav3, red lines to Gadget3, and blue lines to Ramses (reference simulation). One percent agreement (indicated by the grey band) is obtained for* $k \leq 1\ \mathrm{h \cdot Mpc^{-1}}$ *(dashed vertical line), which is typically the scale probed by the Lyα forest. From Schneider et al. (2016).*

clidFlagship simulation (Potter et al., 2017), which has evolved 10^{12} particles in a $(4\ \mathrm{Gpc \cdot h^{-1}})^3$ box with a mass resolution of $\sim 10^9\ \mathrm{M_\odot}$. Nowadays, modern N-body codes all converge below the percent level on the matter power spectrum on scales down to the Mpc. Fig. 5.1 shows the very good agreement on the matter auto-correlation power spectrum between three N-body codes on a wide range of scales and redshifts. For comparison, Fig. 5.2 shows analytical predictions including non-linear corrections of the matter power spectrum compared to a N-body code prediction. The less accurate analytical method (in blue) strongly differs from the numerical prediction (in cyan) for wavenumbers above $0.2\ \mathrm{h \cdot Mpc^{-1}}$. This will result in significantly biased interpretation if it is used as theoretical predictions for cosmological inferences, i.e. the inferred cosmological parameters will not be accurate. The most accurate analytical method (in red) closely follows the numerical prediction. However, this comes with a cost ; in order to be that accurate, the analytical method needs to include many additional nuisance parameters (degrees of freedom) that considerably decrease the precision of the inferred cosmological parameters, i.e. they will have larger uncertainties. This proves the necessity of numerical simulations to fully exploit the cosmological power of cosmological surveys. Indeed, we have now entered the era of high-precision cosmology, with very stringent constraints on cosmological parameters (below the percent level). In the perspective of even better and more data as expected in upcoming projects, theoretical predictions must reach a similar level of precision, in particular in the highly non-linear regime, which is efficiently done with numerical simulations.

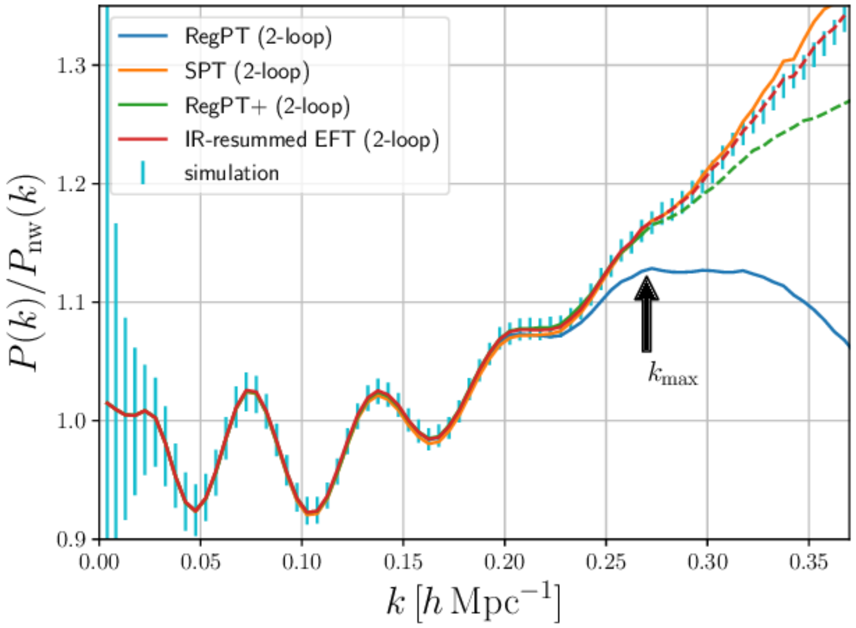

FIGURE 5.2 – *Predictions of power spectrum for different analytical prescriptions. The blue error bar shows the uncertainty of the power spectrum given by the N-body Tree-PM code gadget-2. From Osato et al. (2019).*

5.1.4 Massive neutrinos

The validity of inferred limits on neutrino masses strongly depends on the precise estimation of the impact of their masses on the chosen cosmological observable, in our case the 1D power spectrum of the Lyα forest. In Sec. 1.3.2 of Chapter 1, we derived the impact of massive neutrinos on the linear density fluctuations of cold dark matter and on the power spectrum of the total distribution of matter. In the linear regime, massive neutrinos affect both the cosmic expansion rate and the growth of structures. When neutrinos are still relativistic they behave as additional radiation with a free-streaming length equal to the Hubble radius. When neutrinos become non relativistic during the matter domination era, they act as an additional cold dark matter component. There is a minimum wavenumber above which the free-streaming of massive neutrinos smooth out density fluctuations expressed as,

$$k_{\mathrm{FS}}(z) \approx \frac{0.08}{\sqrt{1+z}} \sqrt{\frac{\Omega_{\mathrm{M}}}{0.3} \frac{\sum m_\nu}{0.1\mathrm{eV}}}\ \mathrm{h}\cdot\mathrm{Mpc}^{-1} \tag{5.12}$$

Broadly,

- for $k > k_{\mathrm{FS}}$: the linear matter power spectrum is suppressed approximately by $\Delta P/P \sim -8 f_\nu$ with weak redshift dependence.
- for $k < k_{\mathrm{FS}}$: the effects of massive neutrinos are negligible.

However, the Lyα forest probes scales in the midly non-linear regime where the suppression is the largest (see Fig. 1.9). Therefore, in order to make reliable predictions on the neutrino masses it is fundamental to have a full non-linear treatement of massive neutrinos.

Neutrinos in the mass range 0.05 eV $\leq \sum m_\nu \leq$ 1.5 eV become non relativistic for redshift between $100 \leq z \leq 300$. Approximately 50% of neutrinos are non relativistic for $z < 200$ and 90% for $z < 100$. Therefore, if we start the simulation at low redshift enough, we can treat all neutrinos as non-relativistic particles. They act as a hot dark matter component, solving the CBE (equation (5.5)) and participating to the gravitational potential (via equation (5.6)), with thermal velocity

$$v_{\mathrm{th}} \sim 150(1+z)\left[\frac{1\mathrm{eV}}{\sum_\nu}\right] \mathrm{km/s}. \tag{5.13}$$

Great caution must be taken when choosing the starting redshift. If the simulation is started too early neutrinos will be treated as massive while still relativistic. Hence, it will amplify their effect and will hinder the inferred mass limit if the simulation is started too late.

Implementing the effects massive neutrinos in cosmological simulations that aim at studying the Lyα forest often come in three approaches ; with rescaling algorithms, with Fourier-space neutrinos or with particle-based neutrinos.

5.1.4.1 Rescaling algorithm

The first approach aims at mimicking the effect of massive neutrinos on the Lyα forest taking advantage of the well-known degeneracy $\sigma_8 - \sum m_\nu$. In Pedersen et al. (2019), the authors show that the effects of massive neutrinos on the growth of structure, within the length scales and redshift range relevant for Lyα forest analysis, are extremely degenerate with a change in σ_8 (or equivalently A the amplitude of primordial perturbations). Therefore, they advise to run additional neutrino-less simulations, only modifying A to cover the neutrino-mass parameter space As such, it does not imply any additional computational cost or memory storage than standard neutrino-less cosmological simulations. However, this degeneracy is not perfect on all scales and redshifts. In particular, it shows discrepancies at 2% on the smallest scales at $z = 3$ when compared with more direct methods implementing neutrino evolution (described in Sec. 5.1.4.3).

5.1.4.2 Fourier-space neutrinos

The Fourier-space method does not properly simulate the evolution of neutrinos but reproduces their impact on the growth of the other components, dark matter and baryons. The local neutrino density is described on a grid and is evolved forward in time using linear theory. The effect of neutrinos on the dark matter can be approximated by adding their gravitational force to that of the dark matter in Fourier space as part of the PM computation (Brandbyge and Hannestad, 2009; Shoji and Komatsu, 2010; Bird et al., 2012; Ali-Haïmoud and Bird, 2013; Archidiacono and Hannestad, 2016; Upadhye et al., 2016; Senatore and Zaldarriaga, 2017; Pedersen et al., 2019).

Such an approach does not require a lot more memory storage (since we do not need to store neutrino positions and velocities) and computational power than simulations without neutrinos. Typically, for a 100 $\mathrm{Mpc} \cdot \mathrm{h}^{-1}$ box evolving collisionless dark matter and collisional gas, adding Fourier-space neutrinos increases by 5% the required computational time. However, because neutrinos are evolved with linear theory, any non-linear evolution in the neutrino component is neglected as well as any gravitational backreaction on dark matter. Indeed, structure growth in the dark matter component induces structure growth in the neutrino component because slowest neutrinos fall into dark matter potential wells. This backreaction effect dragging the neutrinos results from non-linear growth. It is not accounted for in linear theory but the effect is important at the accuracy we wish to achieve with the Lyα forest.

In particular, the authors in Viel et al. (2010) show that the Fourier-space method induces large discrepancies on the Lyα flux power spectrum, at the level of 5% at $z = 0$ and 2% at $z = 4$, when compared with more direct methods implementing non-linear evolution (see next section).

5.1.4.3 Particle-based neutrinos

Particle-based neutrino method is a direct and simple implementation that considers neutrinos as a set of collisionless particles by analogy with N-body methods presented for cold dark matter in Sec. 5.1.3. It takes into account the non-linear evolution of the gravitationally coupled neutrino, dark matter and baryonic components of the matter density. In addition to the bulk velocity given in initial condition by the power spectrum, neutrinos are also given thermal velocities by randomly sampling the Fermi-Dirac distribution. In standard approaches, each neutrino particle is given a random magnitude and direction (Villaescusa-Navarro et al., 2014; Castorina et al., 2014; Massara et al., 2015; Carbone et al., 2016). This is more computationally demanding than Fourier-space neutrinos. Typically, for a 100 $\mathrm{Mpc} \cdot \mathrm{h}^{-1}$ box evolving collisionless dark matter and collisional gas, adding neutrino particles increases by 50% the required computational time (while 5% only for Fourier-based neutrinos).

Unlike cold dark matter, the full 6D phase-space plays a fundamental role in the evolution of neutrino density fluctuations. Because neutrino-particle distribution cannot densely sample the momentum space, thermal motions of neutrinos effectively randomise their position on small scales. At high redshift, neutrinos have high velocities, allowing them to cross multiple times the box such that they quickly lose information about their initial distribution giving rise to significant Poisson noise (Wang and White, 2007). N-body methods rely on the assumption that the set of particles is a faithful representation of the actual phase-space distribution in a given volume. This is the case for cold dark matter with very small velocity dispersion but the free streaming of neutrinos with large thermal velocities makes that the shot noise overcomes the neutrino suppression term. The accuracy can be improved by increasing the number of neutrino particles, with the Poisson contribution to the power spectrum scaling as $\sim k^3 \mathrm{L}_{\mathrm{box}}^3 / N$. This means that the error on the neutrino power spectrum only improves as $1/N$ with the number of particles implemented. To completely remove it only by increasing the number of neutrino particles is out of our computational and memory storage capabilities today. Indeed, we would need about 10^7 times the number of particles we already have. In addition they limit the scalability since fast neutrinos frequently move between processors in parallel code. The simulation TianNu pushed the neutrino problem to extreme scale by simulating $2 \cdot 10^{12}$ neutrino particles, the largest today, in a 1.2 $(\mathrm{Gpc} \cdot \mathrm{h})^3$ box.

Various methods have been proposed to reduce the noise to negligible levels. Some approaches sub-sample the particles to get two independent realizations of the density field. In Viel et al. (2010), the authors disable the short-range gravitational interaction to smooth the gravitational force of neutrinos on small scales. Hybrid methods use the linear treatement for neutrinos at early redshifts with the Fourier-space method, then switch to the N-body treatement once perturbations in the neutrino component become non linear (Brandbyge and Hannestad, 2010). In Banerjee and Dalal (2016), the authors combine particle and fluid descriptions of the thermal species. We can also choose appropriate initial conditions by sampling the Fermi-Dirac distribution, from which the thermal velocities of the neutrinos are drawn, in a regular manner to choose their magnitude and direction, as opposed to randomly sampling it (Banerjee et al., 2018). Finally, a very popular method is the linear-response approximation (Gilbert, 1966), which consists in linearizing the collisionless Boltzmann equation in the gravitational potential (Bond et al., 1980; Ma and Bertschinger, 1994; Singh and Ma, 2003; Ringwald and Wong, 2004; Mummery et al., 2017; McCarthy et al., 2017).

These different approaches all have their advantages and drawbacks and the best implementation depends on the probleme at hand. The authors in (Viel et al., 2010) performed convergence tests on the 1D Lyα power spectrum. They show that at $k \sim 10\ \mathrm{h} \cdot \mathrm{Mpc}^{-1}$, doubling the number of neutrino particles in each dimensions results in a 5-10% effects on the total matter power spectrum and less than 1% on the scales probed by the SDSS Lyα forest ($0.1\ \mathrm{h} \cdot \mathrm{Mpc}^{-1} \leq k \leq 2\ \mathrm{h} \cdot \mathrm{Mpc}^{-1}$), ensuring that numerical errors due to Poisson noise are smaller than observational errors. Increasing the precision of neutrino implementation in numerical simulations is however a major road of improvements for future Lyα surveys such as DESI.

5.1.5 Hydrodynamics of collisional baryonic gas

Dark matter constitutes the backbone of the cosmic structures, where galaxies trace the highest peaks of the dark matter distribution, forming at the center of collapsed halos. However, simulating baryons, even if only making 5% of the energetic budget of the universe, is crucial to make reliable predictions for the Lyα forest. Indeed, the gas density is closely related to that of dark matter on large scales, while on small scales baryonic physics dominates, with for instance thermal broadening and Jeans smoothing. Therefore, the effects of bayronic pressure, non-linear evolution of density perturbations and adiabatic cooling due to the expansion of the universe must be taken into account. Having a full hydrodynamical treatement is mandatory to provide theoretical modeling of the Lyα forest at least at the uncertainty level of observations in order to make strong and reliable scientific interpretations. As an example, the authors in McDonald et al. (2006) used hydrodynamical simulations extended with hydro-particle-mesh (HPM) realizations to perform cosmological analyzes with SDSS Lyα power spectra. However, the HPM simulations exhibit discrepancies in the statistical properties of the Lyα flux distribution at the 20% level when compared with full hydrodynamical simulations (Viel et al., 2006), strongly biasing the inferred scientific interpretation.

In general, the baryonic component is modeled as an inviscid ideal gas, composed of hydrogen and helium. The inviscid feature of astrophysical gas is a fairly good assumption because of its low density. Typically, the density of gas on Earth is 10^{20} H/cc, 1 H/cc in galaxies and 10^{-4} H/cc in the IGM, where 1 H/cc represents one Hydrogen atom per cm^3. Fluid elements in cosmological simulations rarely sample volume smaller than 1 kpc^3, therefore we can safely model baryons as an inviscid ideal gas solving the following set of hydrodynamical equations :

$$\begin{aligned} \text{Mass conservation} \quad & \tfrac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \\ \text{Momentum conservation} \quad & \tfrac{\partial \rho \mathbf{v}}{\partial t} + \nabla \cdot \left(\rho \mathbf{v}^2\right) + \nabla p = \rho \nabla \Phi \\ \text{Energy conservation} \quad & \tfrac{\partial \rho E}{\partial t} + \nabla \cdot [(\rho E + p)\mathbf{v}] = \rho \mathbf{v} \cdot \nabla \Phi . \end{aligned} \tag{5.14}$$

In these equations p is the pressure of the gas, ρ its density, $\mathbf{v}$ its velocity and E its internal energy per unit mass. The gas also experiences gravity and participates in the computation of the gravitational potential in the Poisson equation (equation (5.6)). The set of equation is closed through an equation of state

$$P = A\rho^{\gamma}, \tag{5.15}$$

where A is a constant in space and $\gamma = 5/3$ is the gas adiabatic index.

This set of hydrodynamical equations can be descretized and solved in two general approaches ; we distinguish **Lagrangian methods**, such as implemented in Gadget, where the gas is sampled with fluid particles with their own properties that are followed through space and time, and **Eulerian methods** where the gas is treated as a continuous medium sampled on a grid. Both methods were exploited in this PhD work. The Lagrangian method is used to construct the grids of theoretical predictions covering the cosmological parameter space. The Eulerian method, such as implemented in RAMSES, is used to improve these theoretical predictions that lack baryonic feedbacks and estimate analytical corrections to account for it. This is described in Chapter 6, with the grids of simulations in Sec. 6.1 and the signatures of baryonic feedback in Sec. 6.2 and Sec. 6.3

5.1.5.1 Lagrangian methods

Lagrangian methods, also called Smooth-Particle Hydrodynamics (SPH), are the hydrodynamical analogies of gravity N-body methods. They approximate the continuum dynamics of fluids through the use of sampling particles, which may also be viewed as interpolation points. They follow the equations of motion derived from the hydrodynamical equations. All quantities related to gas particles, such as density or temperature, are computed by smoothing over all neighboring gas particles and the equations are solved for these smoothed quantities. A generic fluid quantity F is computed with

$$F(\mathbf{r}) = \sum_i F(\mathbf{r}_i) W(\mathbf{r} - \mathbf{r}_i, h), \tag{5.16}$$

where W is the smoothing kernel and h is the smoothing length governing the fall of W with respect to the distance between particles. Usually, each individual particle has its own smoothing length so that the total mass enclosed is kept constant. Many smoothing kernels are possible like the gaussian kernel or the cubic spline (Monaghan, 1985). This approach has a very good spatial resolution in high-density regions, when the mean particle distance is small. However this is the opposite when the mean particle distance is large in diffuse regions. This is a major flaw for Lyα forest studies, as we need to control the resolution in the low-density intergalactic medium. Also, shocks are not well captured because of the particle sampling. This requires to add artificial viscosity, for which different forms have been proposed.

5.1.5.2 Eulerian methods

Eulerian methods sample the fluid on a grid that is fixed or adaptative in time. The latter are called adaptative mesh refinement (AMR) methods. It is particulary useful for the Lyα forest because the grid size can be controled to resolve the Jeans length of the IGM. On the opposite of collisionless gas, it is tractable to use grid-based methods for collisionnal gas because we do not need to solve the equations in the 6-dimensional phase-space. Indeed, for collisional gas within a cell (if the cell is small enough), the velocity distribution is Maxwellian and isotropic because it tends to homogenize so that we can neglect the turbulence. Because collisionless gas do not interact, the velocities do not tend to homogeneize. Thus, we need the 3-dimensional velocity space in addition of the 3-dimensional position space.

By discretizing the fluid on a grid, the set of Euler equations can be vectorized and written as one unique equation,

$$\frac{\partial \mathbf{U}}{\partial t} + \nabla \cdot \mathbf{F}(\mathbf{U}) = 0. \tag{5.17}$$

$\mathbf{U} = (\rho, \rho\mathbf{v}, E)$ is the conservative state vector and $\mathbf{F}$ corresponds to the variable flux at the cell boundaries such that temporal variations of $\mathbf{U}$ are directly linked to its spatial variations via the differences in incoming and outcoming fluxes. Given a discretization in 1D space, we use finite difference scheme to write the equation (5.17) as

$$\frac{\mathrm{d}U_i}{\mathrm{d}t} + \frac{1}{x_{i+1/2} - x_{i-1/2}} \left[F(U)^{n+1/2}_{|x_{i+1/2}} - F(U)^{n+1/2}_{|x_{i-1/2}} \right] = 0, \tag{5.18}$$

then,

$$U_i^{t^{n+1}} = U_i^{t^n} + \frac{t^{n+1} - t^n}{x_{i+1/2} - x_{i-1/2}} \left[F(U)^{n+1/2}_{|x_{i+1/2}} - F(U)^{n+1/2}_{|x_{i-1/2}} \right]. \tag{5.19}$$

$U_i^{t^n}$ stands for a numerical approximation to the cell averaged value of cell i at time t^n. $F(U)^{n+1/2}_{|x_{i+1/2}}$ stands for the flux at the cell boundaries between U_i and U_{i+1}, i.e. $U_{i+1/2,\mathrm{left}}$ and $U_{i+1/2,\mathrm{right}}$. Solving the evolution of two constant states separated by a discontinuity is the so-called *Riemann problem*, which must be solved at each interfaces (N+1 Riemann problems to be solved for N cells in 1D). Exact analytical solutions exist but it involves complex non-linear functions and is therefore very costly. Approximate Riemann solvers are more useful (Toro, 2009), e.g. linear, Harten-Lax-von Leer (HLL), HLL Contact (HLLC), HLL with multiple discontinuities (HLLD) or Lax-Friedrich solvers. The choice of the solver is critical because it significantly influences the numerical dissipation of the system. To applicate the Riemann solvers, we have to compute the flux values of each interface of the cell, on the left ($F(U)_{|x_{i+1/2}}$) and on the right ($F(U)_{|x_{i+1/2}}$). There are different interpolation methods to compute it. At the beginning, first-order constant interpolation scheme was used with the piecewise constant method Godunov and Ryabenki (1964). Even if this method is stable it is particularly diffusive and does not capture well shocks. Nowadays, higher-order interpolation schemes are implementation, such as MUSCL (Monotone Upstream-centered Schemes for Conservation Laws) schemes with a piecewise linear interpolation method (van Leer, 1979) or the piecewise parabolic method (PPM) (Colella and Woodward, 1984). They are less diffusive but they require to be coupled to slope limiters to avoid spurious oscillations and keep the interpolation stable. Fig. 5.3 shows the discretization of one of the conservative state vector component in 1D using different interpolation schemes. Fig. 5.4 shows the density solution of the classical Sod shock tube problem with different hydrodynamical solvers implementing different interpolation schemes.

5.1.5.3 Comparison of hydrodynamical implementations

Both approaches, Lagrangian and Eulerian, have been successful at describing the Lyα forest statistical properties, both at low and high redshifts, and they efficiently constrain cosmological and astrophysical parameters. However, it is important to keep in mind their own strengths and weaknesses, with different significance depending on the problem at hand.

The main disadvantage of SPH methods is their inability to control the resolution in diffuse regions. Also, the smoothing length introduced to avoid the discretization noise provokes numerical diffusions. Instead of interacting only with its closest neighbors, SPH particles interact with many particles in their surrounding ($\sim$ 50 particles on average). This can be an advantage if this numerical diffusion is close to be the true physical diffusion, e.g. for conductive media, but

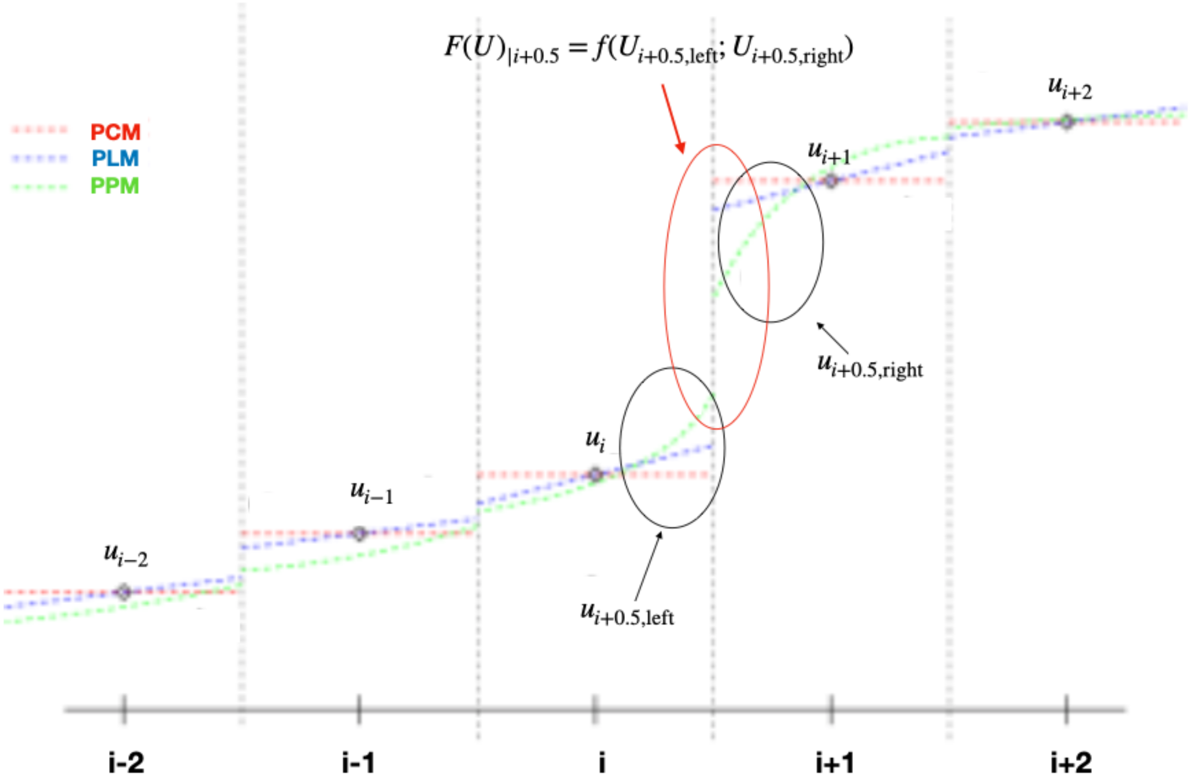

FIGURE 5.3 – *Different interpolation schemes to compute the fluxes at cell boundaries of one of the component of the conservative state vector. The piecewise constant method (PCM) is in red, the piecewise linear method (PLM) is in blue and the piecewise parabolic method (PPM) is in green. The flux values at the right interface of the cell i ($(F(U)_{|i+0.5})$ depends on the field value on the left ($U_{i+0.5,\text{left}}$) and on the right ($U_{i+0.5,\text{right}}$) of this interface, which are interpolated using the different schemes. Adaptated from Colella and Woodward (1984).*

if we consider an insulating medium it significantly changes its physical properties. In the case of the Lyα forest, the numerical diffusion in the low-density IGM is susceptible to exaggerate thermal effects, such as thermal energy injection provoked by AGN feedback.

One of the major argument against Eulerian methods is that they can violate the Galilean invariance of the Euler equations which can make the results sensitive to the presence of large bulk velocities, although the authors in Robertson et al. (2010) argue this is easily overcome with an increase in resolution. Also, numerical viscosity in Eulerian codes will provoke velocity to cancel out in the case of face-on collision when substructures are not resolved enough. Whereas flows with highest resolution would have been able to interpenetrate. Finally, the grids induce privileged directions along the axes, which can have repercussions on collapses of structures and especially on the 1D power spectrum of the Lyα forest which are computed with lines of sight along these axes.

Hybrid methods attempt to find intermediate discretization to combine all the advantages and improve on the weaknesses. For instance, in the moving-mesh code AREPO (Springel, 2011), the mesh is defined as the Voronoi tessellation of a set of discrete mesh-generating points, which are allowed to move freely. This confers this code to have the point of view of a Lagrangian observer so that it is Galilean invariant while retaining the accuracy of mesh-based methods.

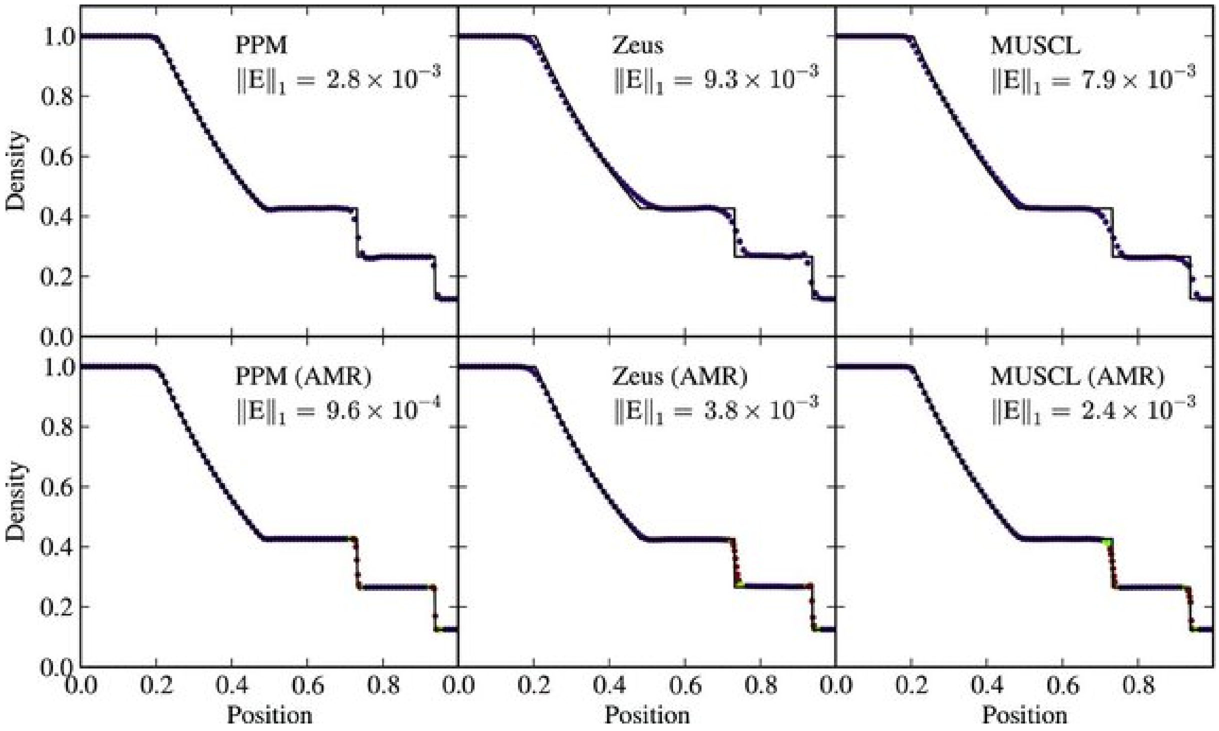

Figure 5.4 – *The density distribution of the classic Sod shock tube for three different solvers (from left to right column) and with (bottomrow) and without (toprow) AMR. In each panel the exact analytical solution of the Riemann problem is shown with the solid line. The integrated absolute deviation from the exact solution is indicated with* $||\mathrm{E}||_1$. *From Bryan et al. (2014).*

In Agertz et al. (2007), the authors performed a comparison between Lagrangian and Eulerian methods on the evolution of a cold gas cloud interacting with an ambient hot moving gas. They show that the smoothing kernel of SPH methods does not allow one to treat well and resolve dynamical instabilities and mixing processes, such as as Kelvin-Helmholtz or Rayleigh-Taylor. Studies with more recent SPH codes have been performed and show that the differences can be reduced but are still there (Beck et al., 2016). However, in practice in cosmological simulations, the scale of such instabilities is not resolved so that the choice of SPH or grid-based methods will not make any fundamental differences on galaxy formation and galaxy evolution. Indeed, the authors from the Aquila comparison Project in Scannapieco et al. (2012) show that there are no systematics biases induced by the choice of the hydrodynamical solver. But the very large dispersion seen in the predictions of galaxy properties is significantly driven by the different sub-grid implementations for stellar and AGN feedback, which will be described in the next section. As such, in the context of cosmological simulations of galaxy formation and evolution, the large differences shown in Agertz et al. (2007) are included in the uncertainties of baryonic feedback implementation since the sub-grid modeling induces large variations in galaxy properties.

5.1.6 Extra baryonic physics

To accurately model the Lyα forest we also need to resolve the formation and evolution of galaxies since they inject considerable amount of energy and strongly alter the distribution of gas in the IGM because of galactic feedbacks (mostly SN and AGN feedbacks). Therefore, it is primordial to produce realistic galaxies because they will heavily influence the dynamics of the Lyα forest. However, to simulate galaxies we need more than the fairly simple adiabatic evolution of the gas modeled by the set of conservative Euler equations. We need to take into account additional astrophysical processes that shape the population of galaxies, such as gas cooling, star formation, baryonic feedbacks etc.. In theory, we add source terms to the set of Euler equations solved by hydrodynamical solvers, which we can re-write ;

$$\begin{aligned} &\text{Mass conservation} && \tfrac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = s_m \\ &\text{Momentum conservation} && \tfrac{\partial \rho \mathbf{v}}{\partial t} + \nabla \cdot (\rho \mathbf{v}^2) + \nabla p = \rho \nabla \mathbf{\Phi} + \mathbf{s_p} \\ &\text{Energy conservation} && \tfrac{\partial \rho E}{\partial t} + \nabla[(\rho E + p)\mathbf{v}] = \rho \mathbf{v} \cdot \nabla \mathbf{\Phi} + s_e. \end{aligned} \tag{5.20}$$

In these equations s_m is the local mass source term caused by star formation, $\mathbf{s_p}$ is the local momentum source term caused by kinetic feedback during SN explosions and s_e is the local energy density source term caused by heating and cooling processes. In the following I will briefly describe the implementation of these processes that are necessary to precisely model the Lyα forest.

5.1.6.1 Heating and cooling processes

Cooling processes are fundamental for the Lyα forest as they lead to the collapse of structures in the IGM. Photoionizing UV background is essential because it ionizes gas that would otherwise cool and collapse into galaxies, regulates the gas accretion of the IGM in massive halos but also drives the IGM ionization state determining its opacity.

The radiative cooling due to atomic processes in the gas, e.g. collisional excitation, recombination or free-free emission, are coupled to the energy equation using tabulated cooling functions $\Lambda_c(\rho, T, Z)$, where Z is the metallicity. These cooling functions typically assume the gas is optically thin, in ionization equilibrium and neglect more than two-body processes.

The UV background comes from quasars and massive young stars and is often modeled as a uniform and isotropic field that evolves with redshift. In practice, it consists in a set of spatially uniform photoionization and photoheating rates of H I, He I and He II. They are often derived from semianalytical synthesis models Haardt and Madau (1996); Faucher-Giguère et al. (2009); Haardt and Madau (2012). However, it has recently been shown that these models are inconsistent during reionization because they lead to different reionization histories than expected and are in disagreement with observational constraints of cosmic reionization. They appear to reionize the universe too early, producing spurious heating of the IGM at early times and consequently affecting its gas pressure smoothing scale that depends on the full thermal history. The authors in Oñorbe et al. (2017) address this issue to self-consistently simulate different reionization models, which efficiently cover the full thermal and ionization histories parameters.

5.1.6.2 Stellar Formation

Cold and dense gas eventually transforms into stars. This results from the very small-scale gravitational collapse of cold molecular gas clouds. The gravitational freefall time, t_{ff}, is inversely proportional to the square root of the density with $t_{\text{ff}} \propto \sqrt{\frac{\pi}{G\rho}}$. Given that the physical scale in cosmological simulations rarely goes below the kpc scale, they are far from resolving the star-formation scale. Thus, to implement this process we rely on **sub-resolution** recipes. The local-mass source term s_m is taken as the opposite of the star-formation rate,

$$s_m = -\frac{\mathrm{d}\rho_*}{\mathrm{d}t}, \tag{5.21}$$

where ρ_* is the local star density. Some gas mass is then converted into new collisionless star particles. Observations support a nearly-universal star-formation efficiency in molecular gas where about 1% of the gas is converted into stars per freefall time. Hence, in the large majority of cosmological simulations, the star-formation rate computation is based on a a Kennicutt-Schmidt type relation (Kennicutt, 1998) as

$$\frac{\mathrm{d}\rho_*}{\mathrm{d}t} = \epsilon\frac{\mathrm{d}\rho}{t_{\mathrm{ff}}}, \tag{5.22}$$

where ϵ is the star formation efficiency, usually around 1%, and ρ the density of the gas. Usually, gas able to convert into stars have, for instance, to be dense enough (Springel and Hernquist, 2003; Agertz et al., 2011; Feldmann et al., 2012), or inside gravitationally bound regions (identified with their virial parameter) (Kuhlen et al., 2012) or prone to gravitational instabilities (identified with their Jeans length) (Oppenheimer et al., 2010; Stinson et al., 2013).

5.1.6.3 Stellar Feedback

Observations show that galaxy formation is a very inefficient process, with only 10% of baryons being locked into galaxies. The peak of galaxy formation efficiency is at 20% in dark matter halos with mass of about $6 \cdot 10^{10}$ $\mathrm{M}_\odot$ and decreases rapidly towards both higher and lower masses (Guo et al., 2010; Moster et al., 2013). Fig. 5.5 shows the observed luminosity function of galactic systems as a function of halo mass for galaxies of the two-degree field galaxy redshift survey in red. The blue solid line shows the predicted luminosity function, taking a constant galaxy-light-to-halo mass ratio, for ΛCDM dark matter halos. The poor match between observations and predictions cannot be alleviated with different galaxy-light-to-halo mass ratios and proves that star formation is inefficient for low- and high-mass halos.

However, the inclusion of cooling processes provokes the *overcooling problem*, where the vast majority of baryons cool down and end up in galaxies and stars, which is in strong disagreement with observations. To avoid this numerical abherration, it is necessary to include stellar feedback, which halts the star-formation efficiency for low-mass galaxies thanks to supernovae-driven winds capable to expel gas from gravitationally-bound systems with low-escape velocities (Larson, 1974; White and Rees, 1978; Dekel and Silk, 1986). The injection of large amount of energy in the ambient medium increases the temperature of the surrounding gas and counteracts the overcooling catastrophe.

Again, cosmological simulations do not reach the scale of star formation and stellar feedback since it is not sufficient to capture the blast wave following the explosion. Thus, we rely on sub-resolution recipes. Assuming an initial mass function (IMF), we can estimate the number of star particles ending as supernovae and compute the total amount of energy (typically $\sim 10^{51}$ erg per supernovae) injected by each star particles. In general, the energy is released thermally, kinetically or both. Early hydrodynamics simulations attempted to inject thermal energy in the nearest gas cells/particles. However, since these cells/particles are also the densest in the simulation and because the cooling rate scales as ρ^2, the net result was an unphysical loss of the supernovae energy due to an artificial excessive cooling tampering the regulation of star formation (Katz, 1992). This is the second overcooling catastrophe.

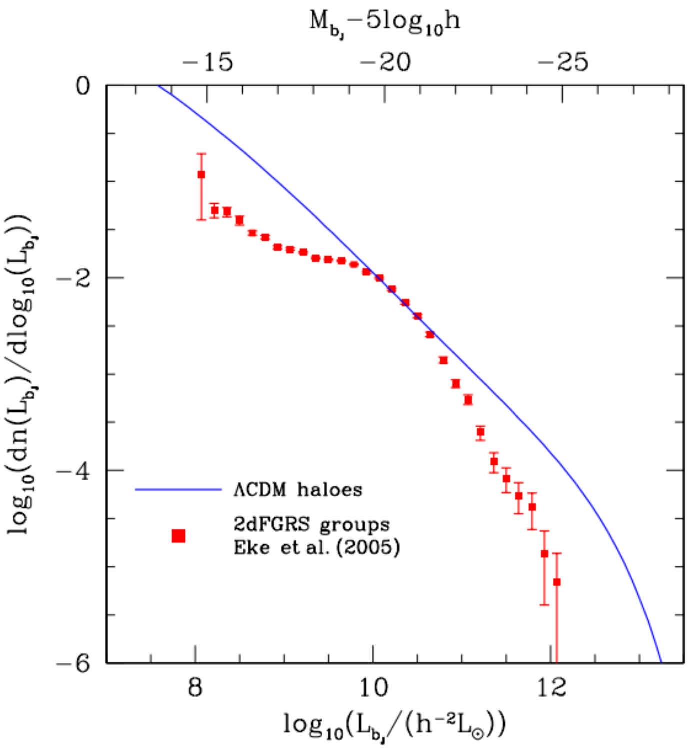

FIGURE 5.5 – *The luminosity function of galactic systems as a function of halo mass. The solid line shows the predictions of the luminosity function taking a constant galaxy-light-to-halo mass ratio (*$\sim 80\ \mathrm{h} \cdot \mathrm{M}_\odot/\mathrm{L}_\odot$*). The red points shows the luminosity function of galaxies from the two-degree field galaxy redshift survey. The y-axis is the galaxy luminosity hosting the halo and the x-axis is proportional to the halo mass. From Baugh (2006)*

To overcome this problem there are many approaches ; some simulations turn off the cooling of the affected gas for a prescribed amount of time ($\sim 10^7$ years) until the SN energy diffuses over a large volume enough (Gerritsen and Icke, 1997; Thacker and Couchman, 2000; Sommer-Larsen et al., 2003; Governato et al., 2007; Agertz et al., 2011). It is also possible to heat the gas probabilistically to reach high enough temperatures (10^6 K) for radiative cooling to become ineffective on time scales of $\sim 10^7$ years (Scannapieco et al., 2008; Teyssier et al., 2013). Another possibility is to include simple models for the multiphase interstellar medium so that the energy is shared between the cold and hot components. The latter is sensitive to the thermal energy and can be effective at reducing the growth of the cold component (Yepes et al., 1997; Springel and Hernquist, 2003; Teyssier et al., 2010).

Stellar feedback recipes still vary widely among different galaxy formation models because they rely on ad-hoc prescriptions. This constitutes a major field of work because identifying the major mechanisms at hand driving the stellar feedback energy would be key elements to improve our understanding of the different galaxy-formation and evolution processes.

5.1.6.4 Supermassive black holes

Fig. 5.5 shows that for high-mass halos, galaxy-formation efficiency is also greatly suppressed. This is not achievable by SN feedback alone without destroying the less massive galaxies, making the predicted faint-to-intermediate galaxy stellar mass function to be in strong disagreement with observations (Oppenheimer et al., 2010; Davé et al., 2011). Therefore, it motivates to consider a completely different feedback process, in particular feedback powered by AGNs, which have been introduced in Sec. 2.1 of Chapter 2.

Observations show that AGNs inject prodigious amount of thermal and kinetic energy in their surrounding medium that couple to the gas by means of relativistic jets, electromagnetic radiations, inflated cavities with strong magnetic fields and non-relativistic outflows. In addition, observations show a tight-correlation between the central black-hole properties and the host galaxies, such as the black-hole-to-stellar mass relation (Laor, 2001; McLure and Dunlop, 2002; Marconi and Hunt, 2003; Häring and Rix, 2004), black-hole mass and stellar-velocity dispersion relation (Ferrarese and Merritt, 2000; Gebhardt et al., 2000; Tremaine et al., 2002), or even correlations between the cosmic star-formation rate (Madau et al., 1996) and quasar luminosity (Boyle and Terlevich, 1998). All evidences point toward the fact that black holes and host galaxy co-evolve and influence each other. Thus, AGN feedback is considered to play a major role in suppressing galaxy formation efficiency for high-mass halos (Tabor and Binney, 1993; Silk and Rees, 1998; Benson et al., 2003; Croton, 2006). In low-mass halos, supernovae winds reduce black-hole growth in the central regions of galaxies by removing cold dense gas until the potential well is deep enough to confine the gas close to the black hole (Dubois et al., 2015; Habouzit et al., 2017).

Supermassive black holes are implemented in numerical simulations with sink particles, which are seed in massive halos, typically with mass above 10^{10} $\mathrm{M}_{\odot}$. These sink particles can accrete gas, often at the Bondi-Hoyle-Lyttleton rate

$$\dot{M}_{\mathrm{BH}} = \frac{4\pi G^2 M_{\mathrm{BH}}^2 \overline{\rho}}{(\overline{c}_s^2 + \overline{u}^2)^{3/2}}, \tag{5.23}$$

with M_{BH} the black-hole mass, $\overline{\rho}$ the mean gas density, $\overline{\mathrm{c}}_{\mathrm{s}}$ the average sound speed and $\overline{\mathrm{u}}$ the average gas velocity relative to the black hole. $\dot{M}_{\mathrm{BH}}$ is limited by the Eddington accretion rate. For accounting for the lack of resolution in accreting regions around black holes and missing the multi-phase structure of the interstellar medium, the accretion rate is often artificially boosted in a density dependent way. The authors in DeGraf et al. (2017) show that introducing stochasitcity in the accretion rate, to mimic episods of clumpy accretion caused by dense-gas clouds could potentially provide the dominant source for black-hole growth of high-redshift galaxies.

In addition of accreting gas, supermassive black holes in simulations can act as AGNs by injecting energy in the ambient medium. AGN feedback can be implemented in two modes : the quasar mode and the radio mode. The quasar mode is believed to happen mostly at high redshifts, when the black hole undergoes fast episodes of growth. It is presumed that the black hole emits large amounts of radiations that heats and ionizes its environment. It is often implemented through energy or momentum injection. The radio mode is believed to happen mostly at low redshifts and accounts for the observed inflated cavities with strong magnetic fields. It is often implemented with mass momentum and kinetic energy injection. The distinction can be done through a threshold value of the accretion rate ; below the threshold the radio mode is triggered, and above the threshold the quasar mode is triggered. However, some simulations do not make any distinction for the different feedback channels asserting that the lack of resolution does not enable to do so. Fig. 5.6 shows a projection of the Horizon-AGN simulation that implements AGN feedback. On the top panels, its temperature map is compared to the one of the Horizon-noAGN simulation that lacks AGN feedback. We can clearly identify hot-gas bubbles due to the injection of thermal energy that ionizes the ambient medium.

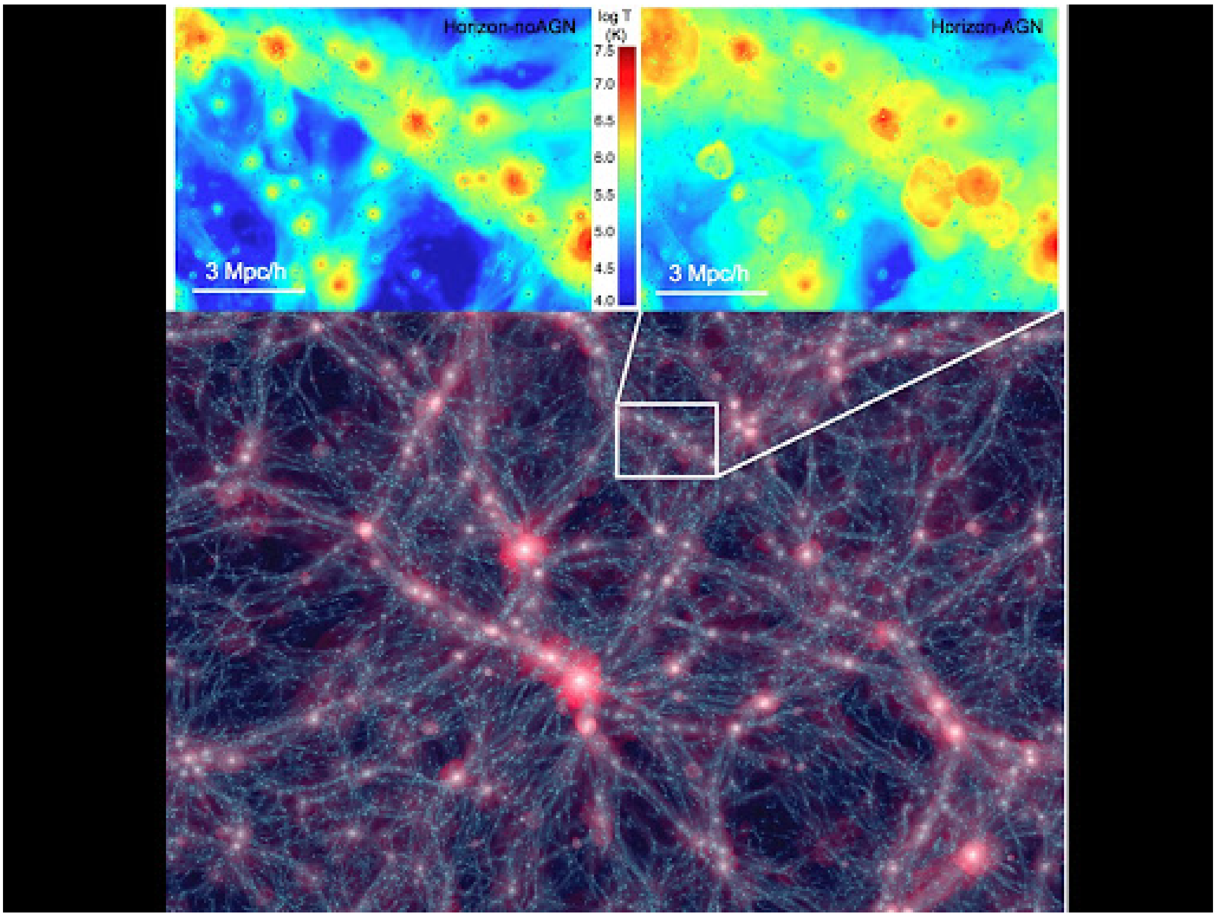

Figure 5.6 – *Projection of the Horizon-AGN simulation at $z = 0$ with th gas density in silver blue and the gas temperature in red. the depth of the projection is* $25\ \mathrm{h} \cdot \mathrm{Mpc}^{-1}$ *and the box extent on the side is* $100\ \mathrm{h} \cdot \mathrm{Mpc}^{-1}$. *The two top panels show the projected temperature maps of the Horizon-noAGN (left) and of the Horizon-AGN simulations. The Horizon-noAGN simulation is implemented with the same physics than Horizon-AGN except AGN feedback. From Dubois et al. (2016)*

The implementation of AGN feedback has been successful at reproducing correlation between black holes and host galaxies (Sijacki et al., 2007; Di Matteo et al., 2008; Booth and Schaye, 2009, 2011; Dubois et al., 2012, 2016; Khandai et al., 2015; Tremmel et al., 2017; Weinberger et al., 2018; Ricarte et al., 2019; Martín-Navarro et al., 2020) and at suppressing star-formation efficiency to avoid the cooling catastrophe in massive galaxies (Puchwein et al., 2008; Khalatyan et al., 2008; McCarthy et al., 2010; Teyssier et al., 2011). Cosmological simulations have proven to be useful tools not only for reproducing correlations between central black holes and host galaxies, but in particular for understanding the mechanisms driving this co-evolution and shaping the population of massive elliptical galaxies. It is now recognized that the growth of black-holes is self-regulated by injecting energy in the medium to suppress cold gas accretion, which in turn suppress star formation. Also, AGN feedback plays a fundamental role at explaining the cosmic evolution and diversity of galaxy morphologies. If elliptical galaxies are known to form from mergers between galaxies that transform disc-type like galaxies to elliptical-like galaxies(Oser et al., 2010; Lackner et al., 2012; Dubois et al., 2014; Kaviraj et al., 2015; Rodriguez-Gomez et al., 2016), AGN feedback enable to perpetuate this transformation by halting in-situ star formation, thus avoiding to re-form disk dominated by rotational support (Dubois et al., 2016).

5.1.6.5 Open issues in cosmological hydrodynamical simulations

Even more physical processes can be included in numerical simulations depending on the problem at hand. For instance, implementing the model equations of radiative transfers is primordial to study the cosmic reionization because radiation alters the thermal, kinetic, and chemical state of the gas. However this is very challenging for simulations on cosmological context because of the high-dimensionality of the problem (frequency and directional dependencies of the photon propagation), and also because the speed of light dramatically increases the required computational power. Magnetic fields, cosmic rays, thermal conduction, metal advection or dust physics can also be modeled. However, the interplay between the large-scale evolution of dark matter, the adiabatic evolution of collisional gas and the inclusion of baryonic feedbacks appear to be sufficient to model the dynamics of the Lyα gas given the already existing uncertainties in modeling baryonic feedbacks and the observational uncertainties of the Lyα forest. Therefore, I did not include further physical processes than those described above in this work.

Simulating realistic galaxies is fundamental to faithfully model the large-scale distribution of matter, and in particular the dynamics of the Lyα forest through the thermal state and gas distribution in the IGM, which is impacted by galaxy formation and evolution through energy injection. State-of-the-art cosmological simulations now reproduce a plethora of observational results and have converged on a wide range of predictions on galaxy properties, gas distribution in the IGM and also statistical distribution of matter on the largest scales.

However, if the basic physical mechanisms required to shape galaxies have been identified, simulations still fail at reproducing observational results simultanesouly in different redshift and mass ranges. Including baryonic feedbacks greatly reduces discrepancies, but numerical simulations still lock too many baryons into stars at $z = 0$. Fig. 5.7 shows the galaxy formation efficiency as a function of halo mass. The solid line is for SDSS/DR7 observational data and the coloured symbols are for different simulations. As in Fig. 5.5, observations show that galaxy formation is a very ineficient process that drops very rapidly towards both higher and lower mass. Only 6% of baryons can condense into stars in massive halos ($\mathrm{M_h} \geq 10^{13}\ \mathrm{M_\odot}$) and around 1% for low-mass halos ($\mathrm{M_h} \leq 10^{10}\ \mathrm{M_\odot}$). The simulation results vary widely but all of them lock too many baryons into stars compared to observations of the real universe. The typical value is around 35%, almost twice as large as required and they all exhibit a quite flat efficiency function.

On top of that, almost all simulations have too high star-formation efficiency at high redshifts ($z \geq 3$), and at the same time, they quench galaxies too early (Fontanot et al., 2009; Weinmann et al., 2012). Therefore, simulated galaxies lack late star formation and most of the stellar mass at $z = 0$ already exists at $z = 2$. This strong early quenching induces a too rapid gas-to-stars conversion leading to discrepancies in gas content of massive galaxies at redshift $z \sim 2$ in simulations (e.g., Illustris-TNG50 (Pillepich et al., 2019) : gas fraction of 10–15%) versus observations (e.g.,(Daddi et al., 2010; Tacconi et al., 2018) : about 50%). In turn, the too-low gas content directly impacts galaxy morphologies because a high fraction of gas is necessary at high redshifts to produce barred-spiral galaxies at $z = 0$ (Martig et al., 2012; Kraljic et al., 2012). This lack of gas is explained by too strong expulsive feedback from young stars and AGN, with mass loading factors in simulations about $\eta \sim 10 - 30$ (Muratov et al., 2015; Nelson et al., 2019) while observations give $\eta \sim 1$ (Schroetter et al., 2019; Förster Schreiber et al., 2019). Lowering AGN and stellar feedback strength could resolve the tension on the gas fraction but it would result in an even more rapid gas-to-star conversion.

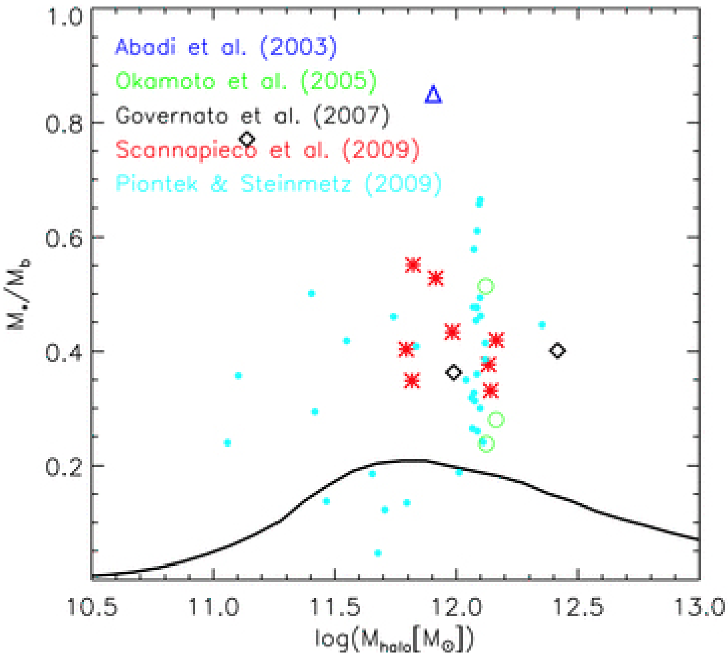

FIGURE 5.7 – *Galaxy formation efficiency as a function of halo mass. The black curve indicates the values required for the* ΛCDM *model to fit the observed SDSS/DR7 stellar mass function. Coloured symbols show the values found for a large number of cosmological simulations. From Guo et al. (2010).*

Conceptually different kinds of solutions have been proposed. For instance, the authors in Gabor and Bournaud (2014) have recourse to coupling between the infalling gas and the accreting disk to halt star formation. The energy conveyed by the accreting gas is converted in turbulent energy, stirs up the disk and lowers high-redshift star-formation efficiency while keeping high gas fraction. Other models invoke the radiation pressure exerted by ionizing photons from massive stars. They argue that the radiation is an important source of pressure in dense, star-forming regions before supernova explosions (Hopkins et al., 2011; Agertz et al., 2013; Ceverino et al., 2014). By modifying the thermal state of the main coolant components, such as the oxygen, it effectively decreases the cooling rates by few orders of magnitude and tampers the star formation efficiency. However, such scenarios were implemented in ideal hydrodynamical simulations lacking the cosmological environment. These require high resolution in the interstellar medium, well below the kpc scale, which is not feasible for large-scale cosmological simulations given the computational power available today.

In summary, simulations struggle at reproducing simultaneously star formation history, galaxy formation efficiency and galaxy morphology. Because simulations are limited by the computational power, they have to rely on sub-resolution recipies that depend on arbitrary parameters. The value of these parameters are either based in physical arguments driven by observations or they are calibrated on observables. Ad-hoc calibrations on one observable create tensions with other observables, as shown for instance with simulations that cannot reconcile at the same time stellar mass at $z = 0$ and gas content at $z = 2$. Therefore, the next goals of computational galaxy formation is to understand which detailed physical processes drive the outcomes of effective physical models. Gaining more physical insights would enable to go beyond the ad-hoc effective models.

5.2 The AMR hydrodynamical code RAMSES

In this section I will describe the Adaptative-Mesh refinement (AMR) hydrodynamical code RAMSES[4] (*Raffinement Adaptatif de Maillage Sans Effort Surhumain*) that I extensively used during my thesis. RAMSES is a massively parallel AMR hydrodynamical code (Teyssier, 2002) that allows solving the Euler equations and gravitational dynamics equations for astrophysical and cosmological systems. Stars and dark matter are modeled by collisionless particles. The collisional baryonic gas is discretized on an adaptatively refined grid and is described by its density, velocity, temperature and energy. RAMSES is written in Fortran 90 and uses the MPI library in order to be run on massively parallel architectures. The developement of RAMSES has been initiated and coordinated by Romain Teyssier and now benefits from a very large international community. This active community has been strongly involved in the development of the initial version by optimizing the code, adding physical processes such as radiative transfers (Rosdahl et al., 2013), supernovae and AGN feedbacks (Dubois et al., 2012), magneto-hydrodynamics (Fromang et al., 2006), and also porting the code on GPUs (Gheller et al., 2015).

The choice of RAMSES for performing realistic simulations of the Lyα forest has been driven by the need of a hydrodynamical code that is performant on a large dynamical range about 10^5. To accurately model the Lyα forest we need a large box, $\sim 100~\mathrm{Mpc} \cdot \mathrm{h}^{-1}$, to simulate the large-scale modes present in the measurements. We also need to go to a resolution down to the kpc scale, i.e. inside galaxies, in order to model baryonic feedbacks that strongly modify the thermal state and distribution of gas in the IGM. Regions that need this resolution only fill a small fraction of the simulation domain volume, less than 1%. Thus, using a uniform grid with $1~\mathrm{kpc} \cdot \mathrm{h}^{-1}$ cell sizes is very inefficient. The AMR feature of RAMSES allows to adapt the resolution as a function of the position on the grid and avoids wasting computational time. However, solving gravitational dynamics, the Euler equations and managing the data structure on a non-cartesian grid implies some difficulties compared to a classical uniform grid. In the following, I will describe the AMR structure in Sec. 5.2.1, the adaptative time step in Sec. 5.2.2, the RAMSES N-body solver in Sec. 5.2.3 and the RAMSES hydrodynamical solver in Sec. 5.2.4.

5.2.1 The AMR structure

AMR methods for hydrodynamical simulations were originally described in Berger and Oliger (1984) and Berger and Colella (1989). It constited in a hierarchy of rectangular grid blocks of various size and resolution, called *nested grids*, which are optimized with respect to the flow geometry. An example of such a **patch-based** structure is shown in Fig. 5.8. The **tree-based** AMR structure was introduced by Khokhlov (1998) with a grid refinement method where the parent cells (also split cells) are refined into child cells (also leaf cells) on a cell-by-cell basis. Thus the resulting data structure is a recursive tree structure allowing to follow more complex flow geometries. However memory and data management are more complex than for patch-based structures, as the former requires tree navigation methods that make cache and memory access difficult to optimize.

The fundamental data structure in RAMSES is called a "Fully Threaded Tree" (FTT), also called and *octree*. The whole simulation box represents the level 0 of refinement. It is initially divided in a regular cartesian grid, called the coarse grid, and corresponds to the level of refinement $\ell_{\mathrm{min}} = 1$. Each cell, or oct, at each level of refinement can be refined by two cells along

---

[4] https ://bitbucket.org/rteyssie/ramses/

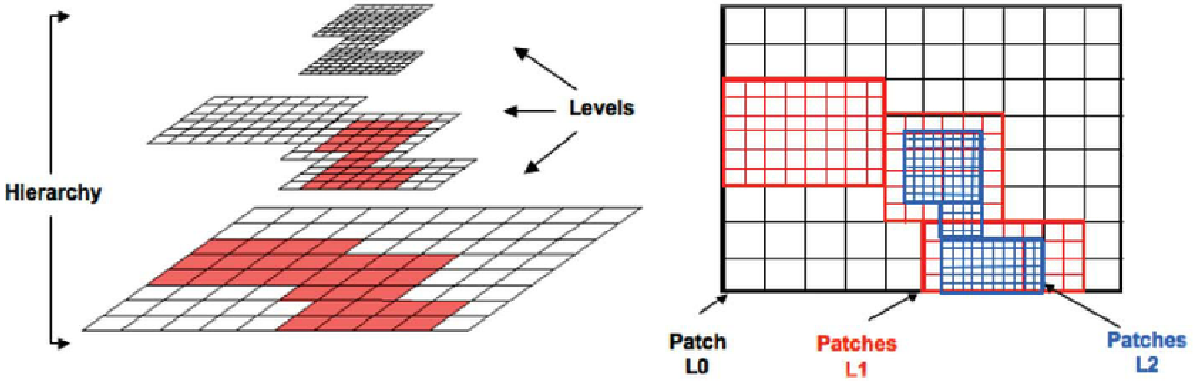

FIGURE 5.8 – *An example of a patch-based structure. Red and blue regions within the initial coarse grid are the more resolved nested grids. From Kamkar et al. (2011).*

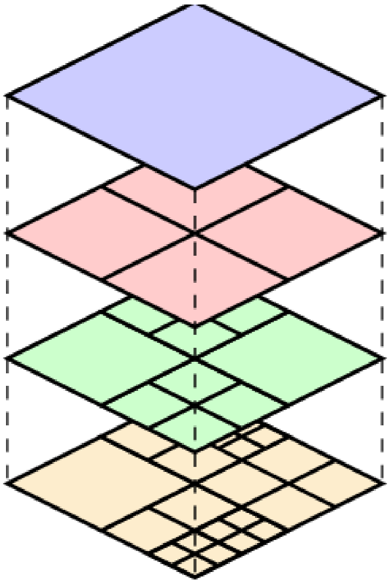

FIGURE 5.9 – *An example of an octree structure in two dimensions. The blue, red, green and yellow grids correspond respectively to the level of refinement $\ell = 0, 1, 2, 3$. The red grid corresponds to the coarse grid. The yellow grid corresponds to the finest grid with the maximum level of refinement, but not all cells are refined up to $\ell = \ell_{\rm max}$. From Drui et al. (2016).*

each dimension (2D : $2^2 = 4$ cells, 3D : $2^3 = 8$ cells). The total effective number of cells on a given level ℓ is $\left(2^\ell\right)^{n_{\rm dim}}$, where $n_{\rm dim}$ is the numer of dimension, but it might not cover the whole volume. A cell at refinement ℓ in a box of size $L_{\rm box}$ has a size $\Delta x_\ell = L_{\rm box}/2^\ell$. Then, the maximal resolution reached is $L_{\rm box}/2^{\ell_{\rm max}}$ where $\ell_{\rm max}$ is the maximum level of refinement. If a cell has no children it is called a leaf cell, otherwise it is called a parent cell. Fig. 5.9 shows an example of an octree structure in two dimensions with three level of refinement.

RAMSES uses a local refinement strategy ; each level ℓ cell can be refined into an oct of $\ell + 1$ cells if a user-defined criterion is met, usually based on the mass of dark matter or baryons contained in the cell. As long as the mass is above a given threshold, the cell is refined until the maximal level is reached. The approximate CPU time overhead for hydrodymical solvers of

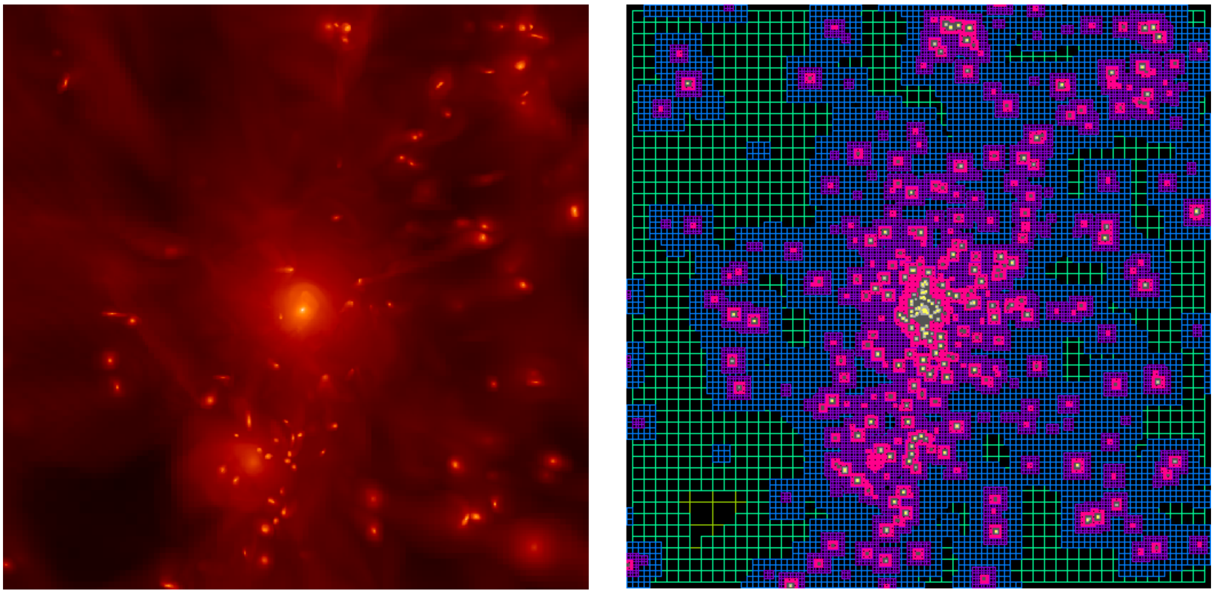

FIGURE 5.10 – *Gas density distribution (left panel) and AMR grid (right panel) of a RAMSES simulation following a galaxy cluster.*

treebased AMR schemes compared to regular cartesian grids is a factor about 2 to 3. Therefore, the refinement strategy of a simulation should be designed to minimize the fraction of the volume to be refined : the maximum of the grid that can be refined should lie in between 30% to 50%.

In addition to the mass criterion, the grid can also be forced to resolve the Jeans length in the gas by at least 4 cells. The authors in Truelove et al. (1997) demonstrated that the Jeans length needs to be resolved by at least a few cells to prevent artificial fragmentation of a self-gravitating gas. But because the grid cannot be refined above $\ell_{\max}$ for obvious computational reasons, we artificially increase the gas temperature by imposing a minimal threshold for the temperature, in order to keep the Jeans length always larger than $4\Delta x_{\min}$, where $\Delta x_{\min}$ is the size of the smallest cell corresponding to $\ell_{\max}$. It can be seen as subgrid model for turubulent motions unresolved in the simulation. In RAMSES this thermal pressure threshold is implemented as a heating function called the Jeans polytrope with $T \propto \rho$.

The AMR structure along with the refinement strategy ensure that the densest regions are evolved at the highest resolution while no computational time is waisted in unnecessarily evolving low-density regions. As an example, Fig. 5.10 shows the gas density and AMR grid for a high-resolution cosmological simulation with RAMSES following a galaxy cluster, where the refinement levels closely follow the gas density distribution.

5.2.2 Adaptative time step control

To ensure numerical stability, it is necessary to prevent gas from traveling more than the length of a cell in one time step. Otherwhise, we can end up with negative mass in a cell, which is obvisouly unphysical. Thus, the time step is determine by the CFL condition (Courant, Friederich and Lewy) (Courant et al., 1967) stating that the time step Δt, the gas velocity v and the size cell must statisfy

$$\frac{v\Delta t}{\Delta x} < C, \tag{5.24}$$

where $C < 1$ is the Courant factor. In RAMSES, C is set to 0.5 such that gas cannot cross more than half a cell.

It appears that the CFL condition is not the same for every level since Δx depends on ℓ where $\Delta x_\ell = 2\Delta x_{\ell+1}$ giving $\Delta t_\ell = \Delta t_{\ell+1}$, where Δt_ℓ is the time step of cells at the level ℓ. The most refined regions need to be updated more often that the coarsest regions. Therefore, in order to avoid wasting computational time updating the whole grid every fine time step, it is adaptative as a function of the level. The *level subcycling* algorithm updates the grids of each level according to its own timestep so that the whole AMR grid is not updated unecessarly too often, and ensures that fine levels statisfy the CFL condition.

Equation (5.24) also shows that resolving dense gas with high velocity significantly slows down the simulation. We easily deduce that simulating AGN and stellar feedbacks is computationally expensive, not only because we need to model small scales, but also because it provokes high-velocity jets, with temperature up to 10^8 K associated to velocity about $10,000\ \mathrm{km}\cdot\mathrm{s}^{-1}$, notably decreasing the physical time step.

5.2.3 N-body solver

To solve the equations of gravitational dynamics (see equation (5.6), equation (5.8) and equation (5.9)), RAMSES implements a N-body PM method on the AMR grid. As explained in Sec. 5.1.3, the first step is to compute the total density fields (i.e. taking into account dark matter, stars and baryonic gas) on the mesh, which is done using a CIC interpolation scheme. The mass m of a particle is distributed to its neighboring cells, with a factor equal to the overlap volume fraction between this cell and a fictuous cell of the same size centered on the particle. This is illustrated in Fig. 5.11.

Then, one needs to solve the gravitational potential Φ on the mesh using the differential form of the Poisson equation. However, this cannot be done using FFTs, as presented in Sec. 5.1.3, since FFTs require regular grids. Also, the use of FFTs requires massive transport of the whole volume data of the simulation, which is difficult to implement and extremely memory intensive when run on distributed memory computers. Another class of Poisson solving schemes are iterative methods, which can be divided in *relaxation methods* (Press et al., 1992) and *Krylov subspace methods* (Saad, 2003). They both start from a first guess Φ_0 for the potential. Relaxation methods, such as Gauss-Seidel or successive over-relaxation (SOR) improve the potential by iteratively damping the residual $r = \nabla^2\Phi - \rho$. Krylov subspace methods, such as the Conjuguate Gradient method, solve the Poisson equation in the form of optimization problem by iteratively minimizing $||\nabla^2\Phi - \rho||^2$. The value of the gravitational potential is then iteratively modified until it converges to a potential which verifies the Poisson equation to a given precision.

The Poisson equation is solved on a level-by-level basis, using a *one-way interface* scheme in which boundary conditions are interpolated from the previous coarser level solution. With these boundary conditions the potential is computed using an iterative method. Originally, a Gauss-Seidel solver was implemented in the initial version of RAMSES. The authors in Guillet and Teyssier (2011) developped an optimized multigrid Conjuguate Gradient method for RAMSES. It is faster to converge since it looks at orthogonal solution at each iteration, but the multigrid approach also allows information from the fine level to flow back to the coarse level. The latter is implemented in the version I used.

Then, the gradients of the gravitational potential are estimated at each grid point by differentiating it with a 5-points finite difference scheme to smooth possible discontinuities at the

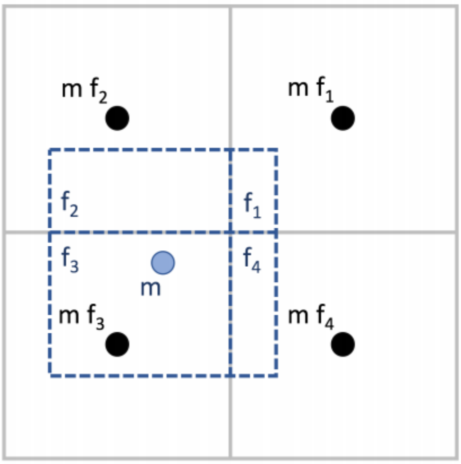

Figure 5.11 – *Two dimensional picture of the CIC method. The mass* m *of the particle is projected onto the grid.* f_i *is the surface fraction (volume fraction in 3D) of the overlap between cell* i *and a fictuous cell centered on the test particle. From Fensch (2017)*

interface between coarse and fine levels. These are interpolated at the particle positions using the same CIC projection kernel. Finally the particle velocities and positions are updated by a second-order predictor-corrector scheme.

5.2.4 Hydrodynamical solver

In RAMSES the Euler equations are solved in their conservative form following the method presented in Sec. 5.1.5.2. Different Riemann solvers are implemented in RAMSES, the simulations I performed were run with a HLLC solver. Interpolations of gas field values across grid cells to estimate fluxes are done with a second-order reconstruction method, the piecewise-linear scheme MUSCL, to limit numerical diffusion, coupled with a slope limitor to avoid unphysical oscillations of the solution near discontinuities.

RAMSES also has many additional physical processes implemented, such as star formation, heating and cooling processes, radiative transfers, magnetic fields or baryonic feedbacks. I will present with more precision in Sec. 6.2 of Chapter 6 the implementation of the processes I used to model the Lyα forest, in particular the subgrid modeling of AGN feedback in RAMSES.

6

Theoretical modeling of the Lyα forest power spectrum

Contents

In the previous Chapter, I described all the numerical tools required to precisely model the Lyα forest with hydrodynamical simulations. I will now detail how I used these tools to construct robust theoretical predictions of the 1D power spectrum of the Lyα forest and the choices that have to been made to concile an efficient exploration of the cosmological parameter space, which implies to run many cosmological simulations, and a modeling accounting for all the necessary physical process that impact our observable, which implies to include computationally expensive sub-grid recipes.

Sec. 6.1 describes the existing grids of simulations covering the parameter space of active neutrinos and thermal warm dark matter. These have been specifically produced for Lyα forest cosmological analyses by the Lyα-group in Saclay with DR9 BOSS data. However, for obvious computational reasons, these simulations lack baryonic physics, which is accounted for with analytical corrections. However these corrections are not at the level of precision of the actual Lyα data. In order to make robust interpretations of these data it is now fundamental to fully and precisely take into account baryonic feedback.

Sec. 6.2 is dedicated to the analysis of the impact of AGN feedback on the 1D power spectrum of the Lyα forest using the Horizon-AGN suite of simulations. The construction and analysis of these simulations have constituted an important part of my PhD.

Finally, in Sec. 6.3, I present the ambitious Extreme-Horizon simulation that I produced. It has been specifically designed to study the impact of the resolution in the IGM and its coupling with baryonic feedbacks. While it has been designed for the Lyα forest, it also brings ground-breaking results for galaxy-formation mechanisms that I will briefly present.

6.1 The grids of simulations

6.1.1 Technical characteristics

In order to explore the cosmological parameter space including the neutrino and thermal warm dark matter sectors, I use three existing grids of cosmological hydrodynamical simulations run with the SPH code Gadget-3. They were designed to vary parameters that efficiently parametrize the 1D power spectrum of the Lyα forest. The goal is to run a maximal number of simulations that probe different dependencies of the power spectrum. They were also constructed to match the requirements imposed by the quality of the BOSS and eBOSS Lyα data.

The first grid explores a flat ΛCDM cosmological model, varying the $\{H_0,\Omega_m,n_s,\sigma_8\}$ parameter space. It is fully described in Borde et al. (2014), I will refer to it as *G_BASE*. The second grid, fully described in Rossi et al. (2014), provides additional simulations for exploring the ΛCDMν model and accounts for dependence of the power spectrum with $\sum m_\nu$. I will refer to it as *G_NU*. Finally, the third grid, described in Baur et al. (2016) provides additional runs for exploring the ΛWDM model and accounts for dependence of the power spectrum with $1/m_X$. I will refer to it as *G_WDM*. In total, the three grids provide N hydrodynamical simulations, exploring the $\{H_0,\Omega_m,n_s,\sigma_8,\sum m_\nu,1/m_X\}$ parameter space allowing to test the dependence of the Lyα power spectrum with this specific set of parameters we want to constrain. The different values of the parameters for each simulation will be presented in Sec. 7.2.1.2.

They are all run with a parallel tree smoothed particle hydrodynamics (tree-SPH) code Gadget-3, an updated version of the public code Gadget-2 (Springel, 2005). The simulations were started at $z = 30$, with initial transfer functions and power spectra computed with

CAMB (Lewis et al., 2000), and initial particle displacements generated with second-order Lagrangian Perturbation Theory. They all include collisionless dark matter, and gas. They use the quick-Lyα option to convert gas particles with overdensities exceeding 10^3 and temperature below 10^5 K into stars. They cover the volume of a periodic 100 $\left(\mathrm{Mpc} \cdot \mathrm{h}^{-1}\right)^3$ box, containing the equivalent of 3072^3 particles of each type. Following a splicing method originally suggested in McDonald (2003), this resolution is obtained by splicing together large-volume and high-resolution simulations, using a transition simulation that corrects the large-box simulation for its lack of coupling between small and large modes, and the high-resolution simulation for its small volume.

In *G_NU*, a particle-type implementation of massive neutrinos is adopted with three degenerate mass species. In Palanque-Delabrouille et al. (2015), the authors show that the 1D power spectrum of the Lyα forest is sensitive to the neutrino mass hierarchy at the level of 0.1% between normal and degenerate hierarchies, and of order 0.3% between inverted and degenerate hierarchies, a level ten times below the simulation statistical uncertainties and almost two orders of magnitude below the data uncertainties. In addition, the authors in Archidiacono et al. (2020) show that changes induced by neutrino mass ordering are so small that even with with very optimistic assumptions about future data changes will remain undetectable. Therefore, the degenerate-mass hypothesis is thus highly justified and we only probe here the sum of the neutrino masses $\sum m_\nu$ and not individual masses.

In *G_WDM*, the impact of thermal relic warm dark matter, or equivalently non-resonantly produced neutrino (see Chapter 1) is mimicked using different transfert functions exhibiting a suppression of power at small scales. In these simulations, the dark matter is fully in the form of warm dark matter. In Baur et al. (2017), the authors implemented simulations with a mix of cold and warm dark matter, which appears to be more in agreement with current observations.

6.1.2 Limitations

Given the large number of required simulations to efficiently cover the cosmological parameter space, the computational power limits the included physical processes. Indeed, these simulations only solve the adiabatic evolution of the baryonic gas but do not include processes for galaxy formation and evolution. Thus, they lack baryonic feedback, which I will show in Sec. 6.2 significantly impact the power spectrum of the Lyα forest by strongly modifying the thermal state and gas distribution in the IGM. I will also show that AGN feedback leads to 2% bias on the scalar spectral index n_s, which represents twice the standard deviations of the current constraints on n_s. Therefore, these cannot be overlooked and it is necessary to have these non-linear effects under control if we aim at making robust interpretation of the data.

We can make the assumption that the impact of baryonic feedback is decorrelated of the cosmological model, which is a fairly good assumption since it corresponds to second order coupling. In this context, we can include the impact of baryonic feedbakcs on the 1D Lyα power spectrum by means of analytical corrections. So far, in Lyα cosmological analysis, analytical corrections from Viel et al. (2013) were used, which I will refere to as V13 henceforth. They were derived by looking at the ratio of the power spectrum with and without AGN and/or stellar feedbacks in hydrodynamical simulations from the OverWhelmingly Large Simulations (OWLS) project (Schaye et al., 2010) run with the three-SPH code Gadget-3. However, while the study explored several scenarios of SN-driven winds, the analysis is based on a specific AGN feedback model because they use one set of parameters for the AGN feedback, coupled with one specific hydrodynamical code used at the OWLS simulation resolution. As I presented in the Chapter 5, galaxy properties in numerical simulations present very large dispertion from one simulation to another driven by the different sub-grid implementations. Indeed, because the sub-grid recipes heavily rely on ad-hoc parametrization, calibration on one observable can create tensions with another observable. In particular, the mass outflow rates in the OWLS simulations, quantifying the amount of gas expelled by galaxies, is about one order of magnitude larger than that of observations. This quantity is of particular interest for the Lyα forest because this hot gas is expelled in the IGM, thus modifying the gas distribution and its thermal state. This too strong feedback is also seen in the mean fraction of galaxies, which is about 15% in OWLS simulations at $z = 2$ when observations give 50%. Hence, results based on these simulations can potentially exaggerate the impact of feedbacks and consequently biase scientific interpretations.

6.2 The impact of AGN feedback with the Horizon-AGN suite

In this section, I extend and improve the analysis done in V13 by giving analytical corrections of the 1D power spectrum of the Lyα forest with uncertainties that encompass the whole range of plausible feedback models in a range of scales and redshifts useful for Lyα surveys, in particular for SDSS and DESI. I led the analysis of this work published in Chabanier et al. (2020) in collaboration with my PhD advisors Frédéric Bournaud, Nathalie Palanque-Delabrouille and collaborators at the Institute of Astrophysics of Paris.

To do so, I use the Horizon-AGN (HAGN) simulation (Dubois et al., 2016) with a more refined resolution in the diffuse IGM that consitutes the Lyα forest gas, which I describe in Sec. 6.2.1.1. I construct a series of additional simulations, presented in Sec. 6.2.1.2 with a set of feedback and feeding parameters spanning the observational uncertainties of galaxy properties at $z = 2$ to estimate the uncertainties related to the feedback model in our corrections. In Sec. 6.2.2 I outline the numerical methods used to derive the flux power spectra from the simulations and I present how well the Lyα forest is reproduced in the Horizon-AGN simulation. I estimate the impact of AGN feedback on the 1D power spectrum in Sec. 6.2.5.1, but I also put an upper and lower bound on the correction to span the whole range of plausible sub-grid parameters in Sec. 6.2.5.2.

6.2.1 The Horizon-AGN suite of simulations

The choice of HAGN as our fiducial simulation was motivated by the following arguments. First, the grid-based method, i.e. Eulerian method, is necessary to control the resolution in the lowest density regions of the IGM that constitute the Lyα forest. Indeed this is not possible with smoothed particle hydrodynamics (SPH), i.e. Lagrangian methods, that spend time evolving the highest density regions. We also need a large enough box, not only not to miss the large-scale modes, but also because of the non-linear coupling of modes during gravitational evolution. Tytler et al. (2009) also state that small boxes are too cold compared to larger boxes because of shock heating being not frequent enough in the small boxes. Lukić et al. (2015) and Borde et al. (2014) used boxes of 80 $\mathrm{Mpc} \cdot \mathrm{h}^{-1}$ to capture all scales when Bolton and Becker (2009) and McDonald (2003) used 40 $\mathrm{Mpc} \cdot \mathrm{h}^{-1}$. With a 100 $\mathrm{Mpc} \cdot \mathrm{h}^{-1}$ box, HAGN is a conservative choice. For the resolution Lukić et al. (2015) requires a 20 $\mathrm{kpc} \cdot \mathrm{h}^{-1}$ cell size for a converged Lyα power spectrum on an uniform mesh without AMR. However, the implementation of baryonic physics with AGN and stellar feedbacks that accurately reproduces properties of galaxy evolution requires a resolution at the kpc scale. Such a dynamical range is computationally too demanding. I will investigate resolution effects on the $P_{\mathrm{Ly}\alpha}$ and AGN correction in Sec. 6.3. Finally, we want to include uncertainties in the feedback model by varying the main sub-grid parameters in order to have $\mathrm{M_{BH}} - \mathrm{M}_*$ and the mean fraction of gas in the range of the observational uncertainties. Therefore our fiducial simulation should be in agreement with observations. HAGN was calibrated at $z = 0$ on the Maggorian relation (Dubois et al., 2012; Volonteri et al., 2016) and appears to reproduce the observed fractions of gas in galaxies at different redshifts, which is one of the main issue in other cosmological hydrodynamical simulations. On the whole, HAGN is a well-suited fiducial simulation because it has the approriate characteristics in terms of box size and resolution to reproduce the Lyα forest and it is in agreement with observational galaxy properties, which is necessary to explore realistic feedback models.

6.2.1.1 The fiducial Horizon-AGN simulations

The cosmological hydrodynamical simulation HAGN is fully described in Dubois et al. (2016), I present in this section the main relevant features. The simulation is run in a box of $\mathrm{L} = 100\,\mathrm{Mpc}\cdot\mathrm{h}^{-1}$. It adopts a classical ΛCDM cosmology with total matter density $\Omega_{\mathrm{m}} = 0.272$, dark energy density $\Omega_\Lambda = 0.728$, amplitude of the matter power spectrum $\sigma_8 = 0.81$, baryon density $\Omega_b = 0.0455$, Hubble constant $\mathrm{H_0} = 70.4$ km $\cdot\mathrm{s}^{-1}\cdot\mathrm{Mpc}^{-1}$, and scalar spectra index $\mathrm{n_s} = 0.967$, compatible with the WMAP 7 cosmology (Komatsu et al., 2011). It contains 1024^3 dark matter (DM) particles, which results in a DM mass resolution $\mathrm{M_{DM,res}} \sim 8\times\,10^7\ \mathrm{M}_\odot$, and initial gas-mass resolution of $\mathrm{M_{gas,res}} \sim 1\times\,10^7\ \mathrm{M}_\odot$. It uses the adaptative mesh refinement code RAMSES (Teyssier, 2002). From the level 10 coarse grid (corresponding to $100\ \mathrm{kpc}\cdot\mathrm{h}^{-1}$ cells), a cell is refined up to an effective resolution $\Delta\mathrm{x} \sim 1\mathrm{kpc}\cdot\mathrm{h}^{-1}$ (level 17 at $z = 0$). Refinement is triggered in a quasi-Lagrangian manner : if the number of DM particles in a cell becomes greater than 8, or if the baryonic mass reaches eight times the initial baryonic mass resolution in a cell, a new level of refinement is triggered.

Gas cooling occurs by means of H and He cooling down to 10^4 K with a contribution from metals following the model from Sutherland and Dopita (1993). Reionization takes place after redshift $\mathrm{z_{reio}} = 10$ due to heating from a uniform UV background from Haardt and Madau (1996). The star formation is modeled with a Schmidt law $\dot{\rho}_* = \epsilon_*\rho/t_{\mathrm{ff}}$ with $\dot{\rho}_*$ the Star Formation Rate (SFR) density, $\epsilon_* = 0.02$ the constant star formation efficiency and t_{ff} the local free-fall time of the gas (Kennicutt, 1998; Krumholz and Tan, 2007). Star formation occurs only in cells with hydrogen gas density ρ exceeding $\rho_0 = 0.1\mathrm{H}\cdot\mathrm{cm}^{-3}$ with a standard 2% efficiency per free-fall time and follow the Schimdt-Kennicutt law (Kennicutt, 1998). Feedback from stellar winds, type Ia and type II supernovae are included to release mass, energy and metals in their surrounding environment assuming a Salpeter Initial Mass Function.

Black holes (BH) are represented by sink particles with an initial mass of $10^5\ \mathrm{M}_\odot$. They can accrete gas in their surrounding environment at the Bondi-Hoyle-Lyttleton rate,

$$\dot{M}_{\mathrm{BH}} = \frac{4\pi\alpha G^2 M_{\mathrm{BH}}^2 \overline{\rho}}{(\overline{c}_s^2 + \overline{u}^2)^{3/2}}, \tag{6.1}$$

with α the dimensionless boost factor, M_{BH} the BH mass, $\overline{\rho}$ the mean gas density, $\overline{\mathrm{c}}_{\mathrm{s}}$ the average sound speed and $\overline{\mathrm{u}}$ the average gas velocity relative to the BH. $\alpha \geq 1$ accounts for the lack of resolution in the accretion disk in star forming gas. We have $\alpha = (\rho/\rho_0)^2$ if $\rho > \rho_0$ and $\alpha = 1$

otherwise. $\dot{M}_{\rm BH}$ is limited by the Eddington accretion rate,

$$\dot{M}_{\rm Edd} = \frac{4\pi G M_{\rm BH} m_{\rm p}}{\epsilon_{\rm r} \sigma_{\rm T} c}, \tag{6.2}$$

with $\sigma_{\rm T}$ the Thomson cross-section, c the speed of light and $\epsilon_{\rm r} = 0.1$ the radiative efficiency. AGN feedback injects a fraction $\epsilon_{\rm f}$ of the radiated energy in the medium in the form of kinetic and thermal energies. It implies,

$$\Delta {\rm E}_{\rm medium} = \epsilon_{\rm f} {\rm L}_{\rm r} \tag{6.3}$$

$$= \epsilon_{\rm f} \epsilon_{\rm r} \dot{M}^2_{\rm BH} c^2, \tag{6.4}$$

where $\rm L_r$ is the radiated energy. The feedbacks come in two modes (Dubois et al., 2012), depending on the value of the ratio of the accretion rate to its Eddington limit

$$\chi = \frac{\dot{M}_{\rm BH}}{\dot{M}_{\rm Edd}}. \tag{6.5}$$

If $\chi > 10^{-2}$ the quasar mode is triggered as it is believed to happen mostly at high redshift, when the BH undergoes fast episode of growth. It is presumed that the BH emits large amounts of radiations that heat and ionize its environment. Therefore thermal energy is injected in a sphere of radius $\rm r_{AGN}$, by increasing the internal energy of the impacted gas cells with $\epsilon_{\rm f} = 0.15$. $\rm r_{AGN}$ is the radius of energy deposition.

If $\chi < 10^{-2}$ the radio mode is triggered. To account for the observed inflated cavities with strong magnetic fields, mass momentum and kinetic energy are injected in bipolar jets with $\epsilon_{\rm f} = 1$. The jets are modeled as a cylindre of radius $\rm r_{AGN}$ and height 2 $\rm r_{AGN}$. $\rm r_{AGN}$ is chosen to be Δx, the size of the smallest cell, after calibration to observations at $z = 0$.

We expect AGN and stellar feedbacks to have different impacts ; SN-driven winds are efficient at expelling gas mostly in low-mass halos because they are not fast enough to overcome the escape velocity of gravitational potentiel of high-mass halos (Dekel and Silk, 1986). AGN feedback are more efficient in high-mass halos because SN winds reduce BH growth in the central regions of galaxies by removing cold dense gas until the potential well is deep enough to confine the gas close to the BH (Dubois et al., 2015; Habouzit et al., 2017). Briefly, we expect a more homogeneous effect for SN feedback compared to AGN feedback. However, if both feedback mechanisms happen at the same time we observe non-linear coupling, in the sense that dense cold gas of SN-driven winds is accelerated by hot outflows, powered by AGN, at much larger scales than without AGN (Biernacki and Teyssier, 2018). The aim of this study is to estimate the impact of AGN feedback on the 1D power spectrum of the Lyα forest ; the feedbacks coupling study is beyond the scope but it needs to be accurately calibrated. The mass loading factor is a key observable for the study of the Lyα forest to validate this coupling calibration because it quantifies the amount of gas expelled by galactic feedback, which ultimately strongly modifies the gas distribution of the IGM. I show in Fig. 6.1 the mass loading factors η in HAGN as a function of of stellar mass at z = 0, z = 1 and z = 2. These have been estimated using outflow rates and star formation rate measurements from Beckmann et al. (2017). These tend to be $\eta \sim 1$ for galaxies with $M_* = 10^{10}$ $\rm M_\odot$ at $z = 1$ and $z = 2$ and tentatively increase toward lower redshifts. This is fully consistent with observations, e.g. see Fig. 7 of Schroetter et al. (2019) or Förster Schreiber et al. (2019).

A companion simulation Horizon-noAGN (HnoAGN) was run without AGN feedback. The same seeds are used so we do not have to account for shot noise.

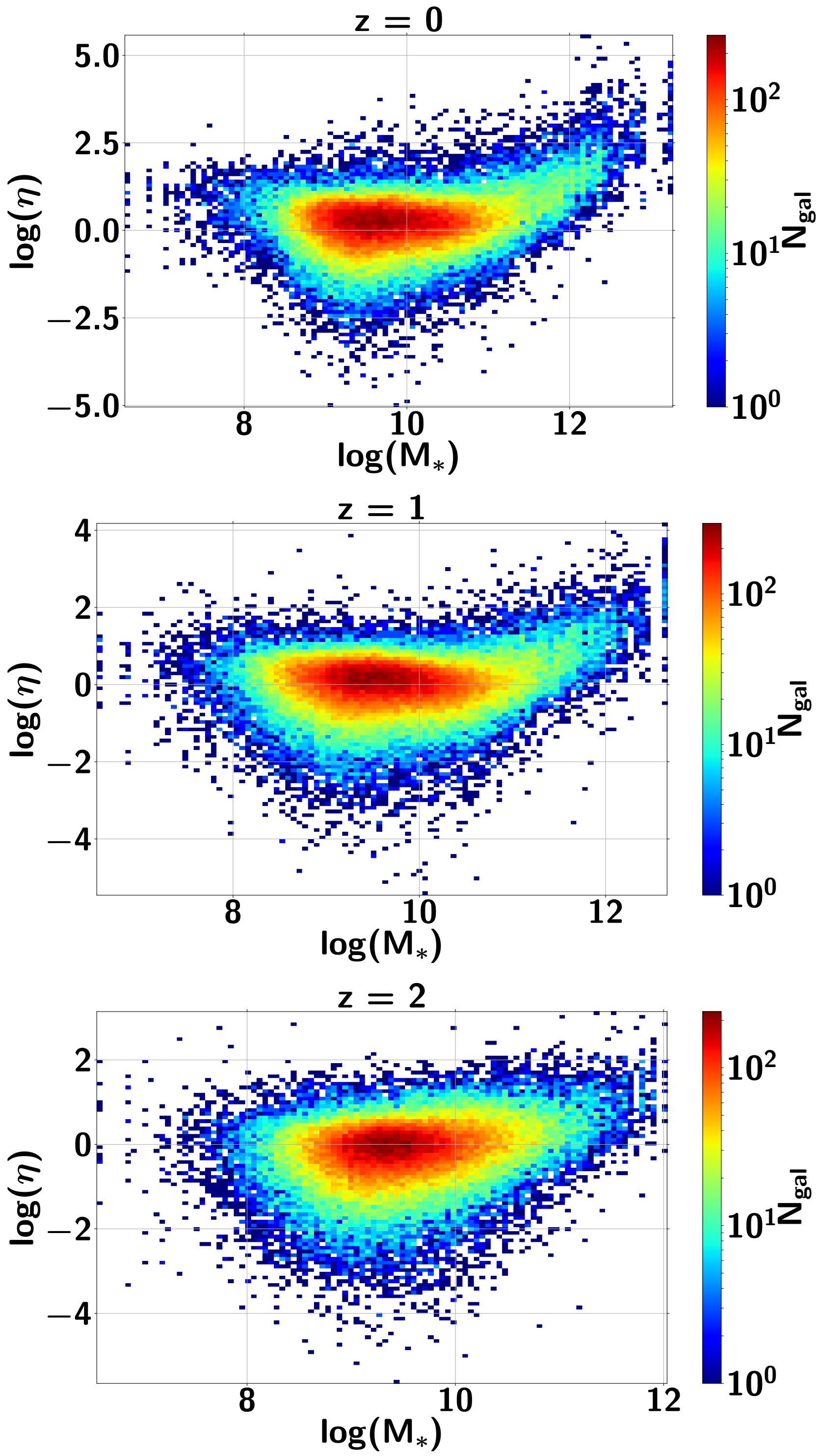

Figure 6.1 – *Mass loading factors of* HAGN, η, *as a function of of stellar mass* M_*, *at* $z = 0$ *(top),* $z = 1$ *(middle) and* $z = 2$ *(bottom). They are defined as* $\eta = \dot{M}_{\rm outflows}/{\rm SFR}$, *where* $\dot{M}_{\rm outflows}$ *is the mass outflow rate and* SFR *is the star formation rate.*

Simulation	α	$\mathrm{r_{AGN}}$	ϵ_{f}
HAGN	$\begin{cases} (\rho/\rho_0)^2 & \text{if } \rho > \rho_0 \\ 1 & \text{otherwhise} \end{cases}$	Δx	$\begin{cases} 0.1 & \text{if } \chi < 10^{-2} \\ 0.15 & \text{if } \chi > 10^{-2} \end{cases}$
HAGNclp10	10% of the time : $10\alpha_{\mathrm{HAGN}}$	$\mathrm{r_{AGN,HAGN}}$	$\epsilon_{\mathrm{f,HAGN}}$
HAGNclp100	1% of the time : $100\alpha_{\mathrm{HAGN}}$	$\mathrm{r_{AGN,HAGN}}$	$\epsilon_{\mathrm{f,HAGN}}$
HAGNr+	α_{HAGN}	$2\Delta x$	$\epsilon_{\mathrm{f,HAGN}}$
HAGNr−	α_{HAGN}	$0.5\Delta x$	$\epsilon_{\mathrm{f,HAGN}}$
HAGNϵ+	α_{HAGN}	$\mathrm{r_{AGN,HAGN}}$	$\begin{cases} 3 & \text{if } \chi < 10^{-2} \\ 0.45 & \text{if } \chi > 10^{-2} \end{cases}$
HAGNϵ−	α_{HAGN}	$\mathrm{r_{AGN,HAGN}}$	$\begin{cases} 0.33 & \text{if } \chi < 10^{-2} \\ 0.05 & \text{if } \chi > 10^{-2} \end{cases}$

TABLE 6.1 – *Summary of the simulations used to estimate corrections and uncertainties due to the AGN feedback model. From left to right, the columns list : simulation name, value of the boost factor, the radius of energy deposition where Δx is the smallest cell, and finally the energy efficiency. I stress that ϵ_{f} can be superior to 1 as it represents the fraction of radiated energy injected in the medium, and not the fraction of total energy.*

6.2.1.2 The set of additional simulations : varying AGN feedback and feeding parameters

To estimate uncertainties that encompasse the whole range of realistic AGN feedback models rather than relying on one single implementation, I performed six restarts from HAGN at redshift 7, when AGN feedback do not have noticeable effects yet. In the six additional simulations, I modify the three main sub-grid feedback and feeding parameters that could impact the Lyα forest :

- **HAGNclp10** and **HAGNclp100** introduce stochasticity in the accretion rate in order to mimic the accretion of massive dense clouds in the interstellar medium, which are not captured by the HAGN resolution. In DeGraf et al. (2017), the authors show that it can impact the evolution of the BH mass at high redshifts. HAGNclp10 has a boost factor and Eddington limit ten times stronger 10% of the time and a hundred time stronger 1% of the time for HAGNclp100. They are run to redshift 2.3 as I will show in Sec. 6.2.5 that α does not impact the corrections above the percent level.
- **HAGNr**+ and **HAGNr**− increase and decrease $\mathrm{r_{AGN}}$ respectively, the other parameters are identical to those of HAGN. They are run to redshift 2.
- **HAGN**ϵ+ and **HAGN**ϵ− increase and decrease ϵ_{f} respectively. They are run to redshift 2.

The sub-grid parameters for each simulation of the suite are summarized in Tab. 6.1. They were chosen so that the $\mathrm{M_{BH}} - \mathrm{M_*}$ relation and the mean fraction of gas $\mathrm{f_{gas}}$ in galaxies span the observational uncertainties.

We take HAGN as the reference at all redshifts for $\mathrm{M_{BH}} - \mathrm{M_*}$ and $\mathrm{f_{gas}}$ for the following reasons. Dubois et al. (2016) calibrated the scaling relation at $\mathrm{z} = 0$ with observations. We also take HAGN as the reference in the scaling relation for highest redshifts given that observational uncertainties are very large, but also because its evolution is weak with increasing redshift following observations analysis (Decarli et al., 2010; Merloni et al., 2010), simulations (Dubois et al., 2012; Sijacki et al., 2015) and analytical models (Croton, 2006). The mean fraction of gas

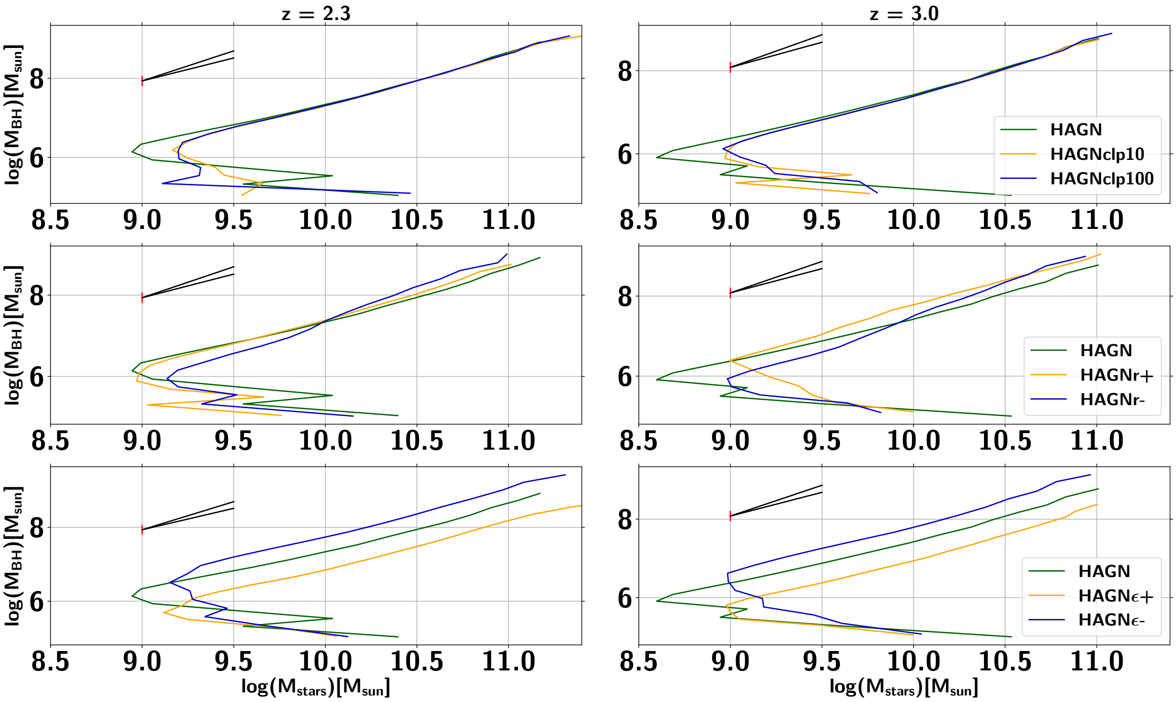

FIGURE 6.2 – $\mathrm{M_{BH}} - \mathrm{M_*}$ *relations for the additional simulations compared to the fiducial simulation* HAGN*. It shows* HAGNclp10 *and* HAGNclp100 *on the first line panels,* HAGNr+ *and* HAGNr− *on the second line panels and* HAGNϵ+ *and* HAGNϵ− *on the third line panels. Left and right panels are at redshift 2.3 and 3.0 respectively. Observation uncertainties are overploted for* $\mathrm{M_*} = 10^{10.5}\mathrm{M_\odot}$*, in red we show* σ_b *the uncertainty on the amplitude, and in black the two extremal slopes authorized by* σ_a

appears to be in the realistic range measured by observations in our range of redshifts, even if no calibration were performed, as shown in Fig. B2 of Welker et al. (2017).

For both relations, we are interested in the deviations to the mean values, hence in the systematic uncertainties more than the intrinsic dispersion. For the mean fraction of gas, Tacconi et al. (2018), give a total observational uncertainty of 0.2. However the intrinsic dispersion largely dominates over the systematic uncertainty and we choose to only take the contribution from this last term, i.e. we take $\sigma_\mathrm{f} = 0.035$. For the scaling relation, we choose to take the global uncertainties from Baron and Ménard (2019) for the slope and amplitude because the systematic term largely overcomes the dispersion. We then have $\sigma_\mathrm{a} = 0.18$ and $\sigma_\mathrm{b} = 0.13$, where σ_a is the uncertainty on the slope and σ_b the uncertainty on the amplitude.

Fig. 6.2 shows the scaling relations for the six additional simulations along with HAGN at redshifts 2.3 and 3.0. We represent the average value of the distribution of the stellar mass for a given bin of BH mass. Observation uncertainties from Baron and Ménard (2019) are overploted for $\mathrm{M_*} = 10^{10.5}\mathrm{M_\odot}$. In red we show σ_b the uncertainty on the amplitude, and in black the two extremal slopes authorized by σ_a. Tab. 6.2 gives the amplitude and slope of the linear fits with deviations to the reference model HAGN in terms of the observational uncertainties. We emphasize that we did the linear fits on the part of the relation where $\log(\mathrm{M_{BH}}) > 7$, as there is not enough statistical power below this value, and on the top branch to be coherent with observations.

Tab. 6.3 gives the mean fraction of gas in the galaxies of the different simulations with deviations compared to HAGN. For the fraction of gas in a galaxy we take the mass of gas

	$a_{2.3}$	$b_{2.3}$	$\Delta a_{2.3}$	$\Delta b_{2.3}$	$a_{3.0}$	$b_{3.0}$	$\Delta a_{3.0}$	$\Delta b_{3.0}$
HAGN	1.34	8.63	0.0	0.0	1.39	8.77	0.0	0.0
HAGNclp10	1.33	8.60	$< \sigma_a$	$< \sigma_b$	1.40	8.76	$< \sigma_a$	$< \sigma_b$
HAGNclp100	1.34	8.62	$< \sigma_a$	$< \sigma_b$	1.43	8.78	$< \sigma_a$	$< \sigma_b$
HAGNr+	1.40	8.76	$< \sigma_a$	σ_b	1.20	8.99	σ_a	$1.7\sigma_b$
HAGNr−	1.59	8.98	$2\sigma_a$	$2.7\sigma_b$	1.62	9.15	$1.7\sigma_a$	$2.9\sigma_b$
HAGNϵ+	1.36	8.17	$< \sigma_a$	$3.5\sigma_b$	1.40	8.32	$< \sigma_a$	$3.5\sigma_b$
HAGNϵ−	1.35	9.04	$< \sigma_a$	$3.15\sigma_b$	1.45	9.22	$< \sigma_a$	$3.5\sigma_b$

Table 6.2 – *Slopes a and amplitudes b of the linear fits for the* $M_{BH} - M_*$ *relation of the additional simulations, such that* $\log\left(\frac{M_{BH}}{M_\odot}\right) = a.\log\left(\frac{M_*}{10^{10.5}M_\odot}\right) + b$. *The second and third columns are at redshift 2.3 and the sixth and seventh columns at redshift 3.0. It gives the deviation to the reference relation of* HAGN *with the* 1σ *deviation being* $\sigma_a = 0.18$ *and* $\sigma_b = 0.13$ *at all redshifts.*

	f_{gas}	Δf_{gas}
HAGN	0.46	0.0
HAGNclp10	0.45	$< \sigma_f$
HAGNclp100	0.42	σ_f
HAGNr+	0.36	$3\sigma_f$
HAGNr−	0.57	$2.7\sigma_f$
HAGNϵ+	0.38	$2.3\sigma_f$
HAGNϵ−	0.56	$2.5\sigma_f$

Table 6.3 – *Mean fraction of gas* f_{gas} *for the resimulations and the fiducial simulation. I also give the deviation to the reference relation of* HAGN *in term of the observational uncertainty* $\sigma_f = 0.035$ *at* $z = 2$.

contained in a cylindre of radius of two times the effective radius and height 2 kpc oriented along the spin of the galaxy, with temperature $T < 10^6$K, to be coherent with the observations, and we compare it to the stellar mass contained in the same volume. We take the mean of all galaxies with mass $10^{10}M_\odot \leq M_{galaxy} \leq 10^{11}M_\odot$ also to be coherent with observations.

We can first notice that the evolution of the scaling relation between $z = 2.3$ et $z = 3$ is weak, as expected, for all the simulations, excepted for HAGNr+ and HAGNr−. On the first row of Fig. 6.2, HAGNclp10 and HAGNclp100 appear to be almost identical to the top branch of HAGN. Tab. 6.2 shows that the slopes and amplitudes are well below the 1σ level at the two redshifts. It confirms the study done in DeGraf et al. (2017) investigating the impact of stochasticity in the BH accretion rates. They show that clumpy accretion is significant at high redshifts for the BH evolution as it enables the accretion rate to outreach the Eddington limit. However they also show that the clumpy and not clumpy accretion models converge around redshift 6. Our simulations confirm that, at our resolution and our redshift range, the clumpy accretion does not have noticeable effect on the Maggorian relation. We only see a $1\sigma_f$ effect for HAGNclp100 on the mean fraction of gas[1].

---

[1] If the stochasticity does not seem to have a strong impact on the mean fraction of gas in galaxies neither on the $M_{BH} - M_*$ relation, it however affects the total mass of the galaxy by decreasing both the mass of gas and stars. We observe galaxies in HAGNclp100 about two times less massive than in HAGN. In that sense, the relations $M_{BH} - M_*$ for HAGNclp10 and HAGNclp100 appear to be at the limit of the observational uncertainty.

HAGNr+ is not very well constrained by the scaling relation, we mostly observe an increase in the fitted amplitude at high redshift. It is in agreement with Dubois et al. (2012), increasing $\mathrm{r_{AGN}}$ leads to higher $\mathrm{M_{BH}}$ because the feedback is less energetic in the medium surrounding the BH. Therefore less gas is ejected and accreted yielding to a lower fraction of gas in galaxies. Indeed Tab. 6.3 shows a mean fraction of gas at $3\sigma_\mathrm{f}$ for HAGNr+. By lowering $\mathrm{r_{AGN}}$ below the cell size, we inject more energy to smaller gas mass which is ejected even further, but less gas is affected. The galaxies in HAGNr− contain more gas than in HAGN at about $2.7\sigma_\mathrm{f}$, because the feedback is stronger than in the first case, to self-regulate its growth the BH accrete less gas.

HAGNϵ+ and HAGNϵ− slopes of the $\mathrm{M_{BH}} - \mathrm{M_*}$ relation are totally in agreement with HAGN at the two considered redshifts. However, the higher the efficiency the less compatible is the fitted amplitude. Our results are again consistent with the study done in Dubois et al. (2012). At a given galaxy mass bin, more massive BHs are obtained if we decrease the efficiency. Following equation (B.15), if the efficiency is decreased, the BH counterbalances by being more massive and accreting more gas in order to inject the same total amount of energy in the medium and self-regulate its growth. The deviation of the amplitude in the scaling relation to HAGN is constant with redshift, and is more than 3σ. The two simulations are less constrained by $\mathrm{f_{gas}}$ with deviations around $2.5\sigma_\mathrm{f}$. Nevertheless, this is consistent with Dubois et al. (2012) as less efficiency in the feedback leads to larger accretion rates hence to less gas in the galaxy.

Our set of additional simulations varying the main feedback parameters appear to largely deviate, i.e. at more than 3σ in terms of observational uncertainty, from at least one of the chosen observables for $\mathrm{r_{AGN}}$ and ϵ_f, and at 1σ for α. By spanning the observational uncertainties we show that we cover the whole range of probable feedback models.

6.2.2 Constructing the power spectrum from simulations

6.2.3 Numerical methods

As described in Chapter 2, the 1D power spectrum of the Lyα forest, $P_{\mathrm{Ly}\alpha}$, is the measure of the variance in the amplitude of the Fourier transform coefficients of the transmitted Lyα flux $\delta_{\mathrm{Ly}\alpha}$ (see equation (2.15) and equation (2.19))

The computation of the transmitted Lyα flux fraction requires the knowledge of the mass, density, temperature at each point of the box. We choose to use SPH equations to perform this 3D mapping for the following reasons. First, it is too strong an assumption to consider that these scalar fields are constant in the AMR cells. Then, the state of an AMR cell influences its neighbors. We want to parametrize the fields in the box as smooth functions and not as unrealistic step-functions. To do so, we transform the AMR gas cells into particles. We loop over each gas cell and place a particle with the total mass of the cell at its center using the rdramses tool[2]. We use the 3D cubic spline kernel introduced in Monaghan and Lattanzio (1985) to smoothly distribute the quantities of interests of each particles over its neighboring cells :

$$W\left(q_j\right)=\begin{cases}\left[1+q_j^2\left(-1.5+0.75q_j\right)\right]\cdot\frac{1}{\pi} & \left|q_j\right|\leq 1\\ \left[0.25\left(2-q_j\right)^3\right]\cdot\frac{1}{\pi} & 1<\left|q_j\right|\leq 2\\ 0 & \left|q_j\right|\geq 2\end{cases}\tag{6.6}$$

where $q_j=\left|\mathbf{r}-\mathbf{r}_\mathrm{j}\right|/h_j$ is the reduced distance to particle i. The smoothing length h is chosen such that the volume inside the sphere of radius h is equal to the volume within the considered

---

[2] `http://www.astro.lu.se/~florent/rdramses.php`

cubic cell. Then,

$$\left[\frac{\mathrm{L_{box}}}{2^{\ell}}\right]^3 = \frac{4}{3} \times \Pi \times \mathrm{h}^3, \tag{6.7}$$

with ℓ being the level of the cell. The simulation does not go below level 15 for redshift above 2. Finally, we can derive each scalar fields at every points of the box using the following SPH equation :

$$A(\mathbf{r}) = \sum_j m_j \frac{A_j}{\mathrm{N_{H0,j}}} W\left(|\mathbf{r} - \mathbf{r}_j|, h_j\right), \tag{6.8}$$

where A one of the scalar quantity, $\mathbf{r}$ a position in the cube, h the smoothing length and W the kernel functions described in equation (6.6), and finally $\mathrm{N_{H0}} = \frac{n_{H0}}{n_H}$ is the neutral fraction of hydrogen. The index j loops over all the gas particles in the simulation box.

The neutral hydrogen fraction is fundamental for the computation of $P_{\mathrm{Ly}\alpha}$. To model the chemistry of the gas, we consider the IGM as having the primordial gas abundances with hydrogen abundance $X = 0.76$, and helium abundance $Y = 0.24$. This is in agreement with the recent CMB observations (Planck Collaboration et al., 2018a). $\mathrm{N_{H0}}$ is estimated following the classical hypothesis that the IGM gas is optically thin and in ionization equilibrium but not in thermal equilibrium. We only consider collisional ionization cooling, radiative recombination cooling and photo-heating from a uniform UV background to impact the chemical evolution of the 6 atomic species $\mathrm{H_0}$, $\mathrm{H^+}$, He0, $\mathrm{He^+}$, $\mathrm{He^{2+}}$ and $\mathrm{e^-}$. It leads to the following set of equations :

$$\mathrm{n_{H0}} = \mathrm{n_H}\alpha_{\mathrm{H^+}}/(\alpha_{\mathrm{H^+}} + \Gamma_{\mathrm{e,H0}} + \frac{\Gamma_{\gamma,\mathrm{H0}}}{\mathrm{n_e}}) \tag{6.9}$$

$$\mathrm{n_{H^+}} = \mathrm{n_e} - \mathrm{n_{H0}} \tag{6.10}$$

$$\mathrm{n_{He^+}} = (\mathrm{n_{He0}}\alpha_{\mathrm{He^+}})/(\Gamma_{\mathrm{e,He0}} + \frac{\Gamma_{\gamma,\mathrm{He0}}}{\mathrm{n_e}}) \tag{6.11}$$

$$\mathrm{n_{He^+}} = \mathrm{Y}\mathrm{n_H}/(1 + \frac{\alpha_{\mathrm{He^+}}}{\Gamma_{\mathrm{e,He0}} + \frac{\Gamma_{\gamma,\mathrm{He0}}}{\mathrm{n_e}}} + \frac{\Gamma_{\mathrm{e,He^+}} + \frac{\Gamma_{\gamma,\mathrm{He^+}}}{\mathrm{n_e}}}{\alpha_{\mathrm{He^{2+}}}}) \tag{6.12}$$

$$\mathrm{n_{He^{2+}}} = \mathrm{n_{He^+}}(\Gamma_{\mathrm{e,He^+}} + \frac{\Gamma_{\gamma,\mathrm{He^+}}}{\mathrm{n_e}})/(\alpha_{\mathrm{He^{2+}}}) \tag{6.13}$$

$$\mathrm{n_e} = \mathrm{n_{H^+}} + \mathrm{n_{He^+}} + \mathrm{n_{He^{2+}}}, \tag{6.14}$$

$$\tag{6.15}$$

with α the recombination rate, Γ_e the collisional cooling rates and Γ_γ the photoionization rates. If we consider that helium is fully ionized either once or twice, and if we neglect the other ionization state, then the electron fraction is only function of the hydrogen density and the neutral hydrogen fraction from equation (6.15) can be easily computed. We have $\mathrm{n_e} \sim 1.15\mathrm{n_H}$ and $\mathrm{n_e} \sim 1.10\mathrm{n_H}$ for the full first and second ionization respectively. We checked that making the assumption that Helium is either once or twice ionized does not significantly change the corrections with differences at the level of 10^{-3}. In the following we make the calculations of the free electron fraction considering that Helium is fully ionized twice. The set of equation reduces to

$$\mathrm{n_{H0}} = \mathrm{n_H}\alpha_{\mathrm{H^+}}/(\alpha_{\mathrm{H^+}} + \Gamma_{\mathrm{e,H0}} + \frac{\Gamma_{\gamma,\mathrm{H0}}}{1.15\mathrm{n_H}}). \tag{6.16}$$

We use the radiative cooling rates from Abel et al. (1997), the collisional cooling rates from Katz et al. (1996) and the photoionization rates from Theuns et al. (1998).

Once all the required fields are computed for each gas particle, we extract 50,000 LOS parallel to one of the axis of the box (which is not the same for all LOS), and whose origin and axis are randomly drawn, following the traditional procedure (Croft et al., 2002; Gnedin and Hamilton, 2002).

We divide the spectra in $\mathrm{N_{bin}} = 2048$ bins with coordinate $x(\mathrm{Mpc} \cdot \mathrm{h}^{-1})$ in real space and $u(\mathrm{km} \cdot \mathrm{s}^{-1})$ in velocity space, such that $u(1+z) = xH(z)$. For each pixel j of each LOS, we use the SPH equation of equation (6.8) to derive the density $\mathrm{n_{H,j}}$, the temperature T_j and the peculiar velocity v_j of the gas in this pixel. The observed velocity is then,

$$v_{obs,j} = v_{j||} + u_j = v_{j||} + \frac{H(z)}{1+z}x, \tag{6.17}$$

where $v_{j||}$ is the peculiar velocity of the gas along the LOS and x the pixel coordinate in real space. From this we estimate the optical depth τ for H0 using an analytic approximation to the Voigt-Hjerting function, with which Voigt-profile are modeled following Tepper-García (2006) as described in Chapter 2. In velocity space, peculiar velocities modify the optical depth by shifting the absorption positions and broadening the lines (McDonald, 2003) (see Fig.1 of Lukić et al. (2015)). We thus have

$$\tau_s(u) = \int_0^{L/2} \tau(x') \frac{1}{\sqrt{2\pi} b(x')} \exp\left(-\left(\frac{u - v_{obs}(x')}{b(x')} \right)^2 \right) dx', \tag{6.18}$$

where $\tau_s(u)$ is the optical depth in redshift space at velocity coordinate u, $\tau(x')$ the optical depth in real space at spatial coordinate x' and $b(x') = \sqrt{2k_B T(x')/m_H}$ is the Doppler parameter with k_B the Boltzmann constant and m_H the mass of the hydrogen atom. All $P_{\mathrm{Ly}\alpha}$ computations in the following are done in redshift space.

We highlight the fact that, on the contrary of most of the hydrodynamical simulations working with the Lyα forest, we do not rescale the optical depths such that the mean flux $\langle \phi(z) \rangle$ match the observations. We are interested in differences due to AGN feedback and it can include differences in the mean flux. Moreover there are no reasons that HAGN and HnoAGN should have the same mean flux as they do not represent the same universe. Then we can compute at each pixel j the Lyα flux density contrast $\delta_{Ly\alpha,j}$ where the flux is $\phi_j = e^{-\tau_j}$, and the mean flux $\langle \phi(z) \rangle$ is estimated from the ensemble of the pixels along all LOS. Finally, the 1D power spectrum, $P_{\mathrm{Ly}\alpha}$ is constructed by taking the Fourier Transform of the transmitted flux fraction field using a Fast Fourier Transform (FFT) algorithm.

6.2.4 The Lyα forest in Horizon-AGN

I show in Fig. 6.3 the evolution with redshift of the $P_{\mathrm{Ly}\alpha}$ for HAGN in plain lines. The error bars represent the root-mean-square of the 50,000 LOS sample, they are well below the percent level at all redshifts. The yellow stars are the BOSS/eBOSS $P_{\mathrm{Ly}\alpha}$ data points derived in Chapter 4 at redshift 3, we include statistical and systematics uncertainties in the error bars. Simulations and observations are in broad agreement, both in shape and amplitude. We do not require better agreement, as we are only interested in differences produced by AGN feedback.

Fig. 6.4 presents the mass-weighted temperature-density diagram of HAGN and HnoAGN at redshift $z = 2$ in logarithmic scales. The four populations constituing the baryonic gas are clearly visible ; the cold diffuse density IGM, the hot IGM, the hot high density virialized gas from clusters and finally, the cold condensed star forming gas. The cold IGM phase constitutes the Lyα forest we are interested in, it contains a very large fraction of the baryonic gas, both in volume and mass, and follows a linear relation between log(T) and log(ρ), as seen in observations and other cosmological hydrodynamical simulations including cooling (Borde et al., 2014; Lukić et al.,

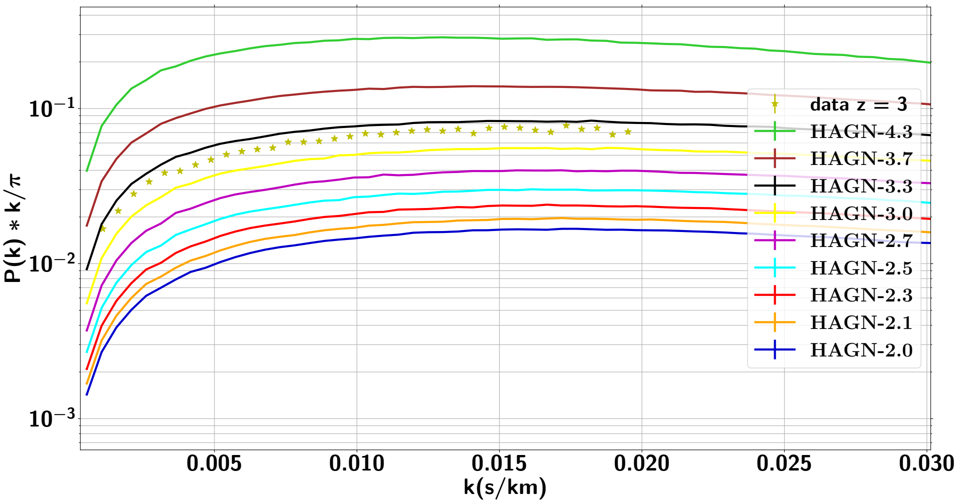

FIGURE 6.3 – *Comparison of the* $P_{\mathrm{Ly}\alpha}$ *of* HAGN *at different redshifts in plain lines with the eBOSS DR14 data at redshift 3.0 for the yellow stars. Error bars represent the statistical error on the 50,000 LOS for the simulation* $P_{\mathrm{Ly}\alpha}$*, and the combination of statistical and systematics uncertainties derived in Chapter 4 for the observational data.*

2015). The Jeans polytrope is also clearly visible at high densities, for $\rho > \rho_0 = 0.1\mathrm{H}\cdot\mathrm{cm}^{-3}$, with the following EoS,

$$\mathrm{T} = \mathrm{T}_0\left(\frac{\rho}{\rho_0}\right)^{\mathrm{p}-1}, \tag{6.19}$$

where $p = 4/3$ is the polytropic index of the gas. The reason to artificially increase the temperature of condensed star forming gas in HAGN is twofold. First, to increase the Jeans' lenght and avoid numerical artificial instabilities (Truelove et al., 1997) as presented in Chapter 5, but also to account for the thermal heating of the ISM by SNe explosions (Springel et al., 2005). However, as the neutral fraction $\mathrm{N}_{\mathrm{H}0}$ is greatly dependent on the temperature so is the $P_{\mathrm{Ly}\alpha}$. Hence, we checked that taking $\mathrm{T} = 10^4\mathrm{K}$ in post-processing instead of keeping the artificially enhanced temperature from the simulation, for all gas cells with $\rho > \rho_0$ do not change the results. Of course, modifying the temperature in such dense regions does not impact the $P_{\mathrm{Ly}\alpha}$ that dominantly probes the very diffuse gas. I will come back to the comparaisons of the two diagrams in Sec. 6.2.5

6.2.5 Results

6.2.5.1 Impact of AGN feedback on the $P_{\mathrm{Ly}\alpha}$

Fig. 6.5 shows the corrections β estimated from the fiducial simulation HAGN. We take the correction due to AGN feedback as the deviation to one of the ratio of the $P_{\mathrm{Ly}\alpha}$ in HAGN to those in HnoAGN using the same 50,000 LOS, such that,

$$\frac{\mathrm{P}_{\mathrm{Ly}\alpha}(\mathrm{HAGN})}{\mathrm{P}_{\mathrm{Ly}\alpha}(\mathrm{HnoAGN})} = 1 + \beta. \tag{6.20}$$

The results are displayed at different redshifts from $z = 4.25$ to $z = 2.0$. We observe a suppression of power that increases with decreasing redshifts and increasing scales. The enhancement of

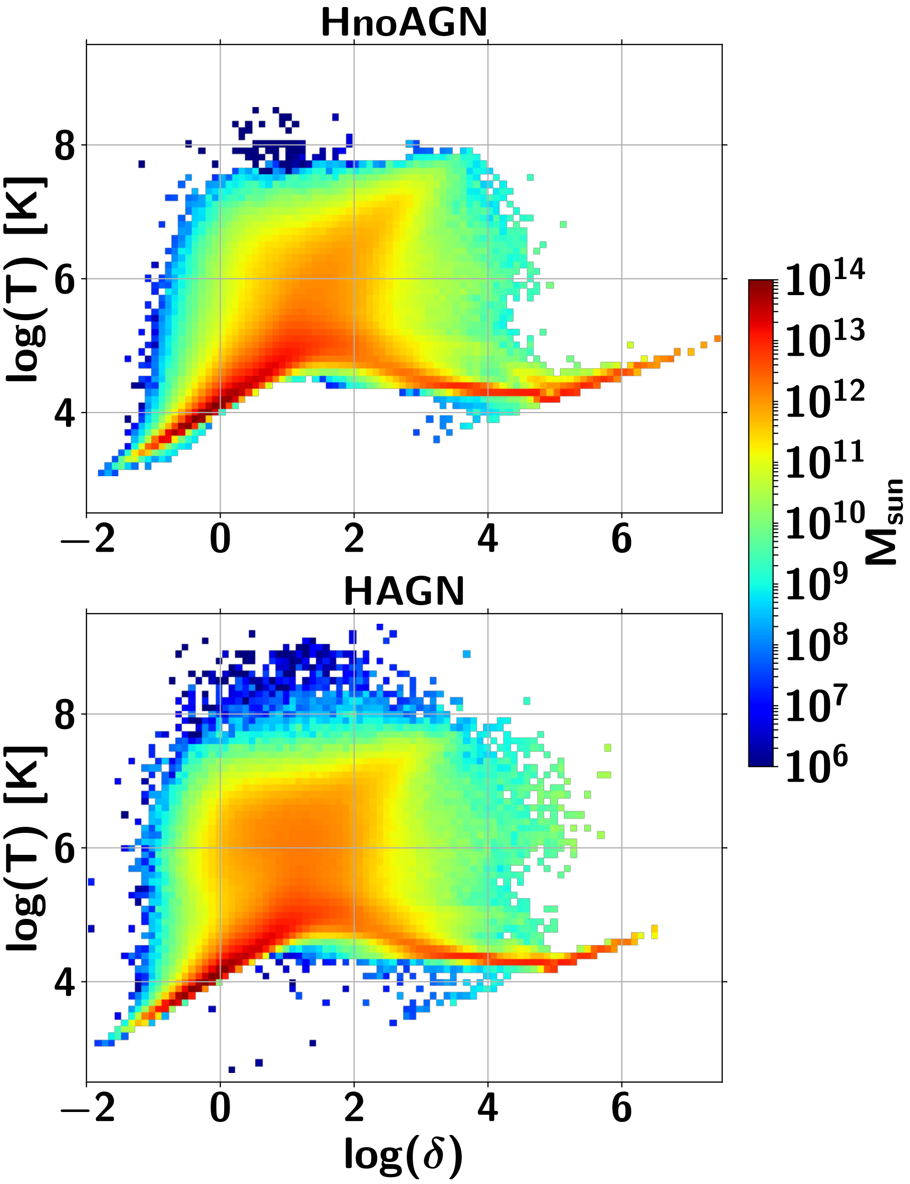

FIGURE 6.4 – *Mass-weighted temperature-density diagrams of the gas cells in* HAGN *(bottom) and* HnoAGN *(top) at redshift 2.0 in logarithmic scales. The temperature T is in Kelvin, and δ is the density contrast $\delta = \rho/\overline{\rho} - 1$.*

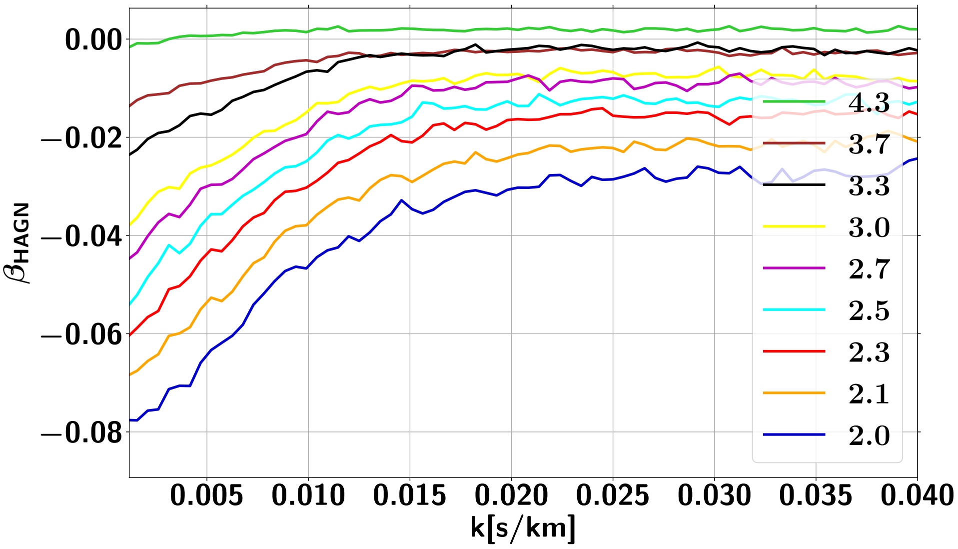

FIGURE 6.5 – *Corrections* β_{HAGN} *of AGN feedback using the fiducial simulation* HAGN. *The different lines are for the nine different redshifts from* $z = 4.25$ *to* $z = 2$.

suppression of power at large scales is already noticeable at $z = 4.25$ and rises from less than 1% to 8% at $z = 2.0$.

As previously said in Sec. 6.2.2 we do not rescale the mean flux for HAGN and HnoAGN and we observe a global decrease of power, i.e. an increase of the mean flux, with AGN feedback. As shown in Fig. 6.6, we observe a strong decrease in the number of pixels with low-flux transmittivity. It reflects the combination of a net increase of temperature, ionizing the ambient medium, but also the redistribution of gas from small to large scales. The gas heating is clearly visible on the projected temperature maps of HAGN and HnoAGN in the top panels of Fig. 6.7. The left panel (HAGN) displays hotter bubbles than the right panel (HnoAGN), that extend to larger scales and reach the IGM. This is in agreement with the temperature-density diagrams of Fig. 6.4 where we observe more pixels in the diffuse region, i.e. $\log(\delta) < 2$ with temperature $T > 10^5$K in addition of the appearance of pixels with $T > 10^8$K with AGN feedback. Indeed the hot IGM contains 18% of the mass in HAGN and 12% in HnoAGN. The temperature is also higher in the dense region, i.e. $\log(\delta) > 2$, but the heating is less efficient as the temperatures do not go above $10^7/10^8$K. This net increase of temperature is due to the injected thermal energy of quasar mode black holes that dominates compared to radio mode in our redshift range and ionizes the surrounding gas. Because the $P_{\text{Ly}\alpha}$ probes neutral hydrogen and because there is more ionized gas the power spectrum exhibits a suppression of power at all scales.

The redistribution of gas is subtle but distinguishable on the projected density maps of HAGN and HnoAGN on the bottom panels of Fig. 6.7. Dense gas bubbles around dark matter halos are less confined and spread to larger radius in HAGN compared to HnoAGN. It is clearly visible on the zoom figures of Fig. 6.8 described later. On the temperature-density diagrams of Fig. 6.4, the under-dense region, i.e. $\log(\delta) < 2$, is more populated in the AGN feedback case, to the detriment of the dense region, i.e. $\log(\delta) > 2$. Indeed, The hot and cold IGM contain 86% of the mass versus 82% of the mass for HAGN and HnoAGN respectively.

To disentangle the gas heating and mass redistribution effects, we compute the same correction where the heating is switched-off in HAGN. To do so, we impose the HAGN temperature-

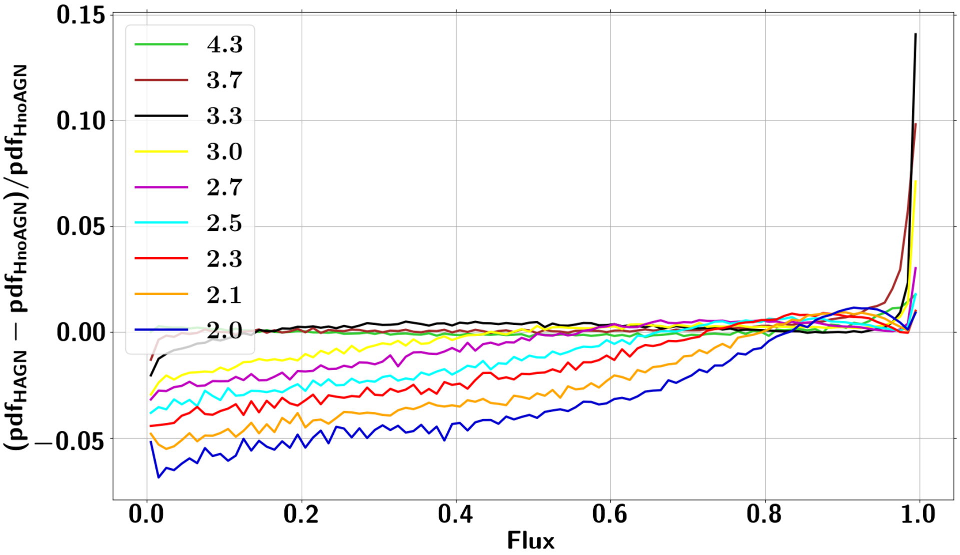

FIGURE 6.6 – *Differences in the flux probability distribution betwen* HAGN *and* HnoAGN *at all redshifts using pixels from the 50,000 LOS sample.*

density diagram to be the same than HnoAGN. We estimate the probability distribution functions (PDF) of the temperature in 100 density bins in HnoAGN. Then, for each gas particle we draw a temperature from the appropriate temperature PDF depending on the gas density. We introduce noise by decorrelating the temperature at the very small scales, therefore we also apply this modification of temperature in HnoAGN. Fig. 6.9 shows this correction 'without heating' from AGN feedback. The increase of power on large scales is coherent with the study from Chisari et al. (2018). AGN feedback redistributes gas from the small scales to the large scales, hence the matter power spectra show a suppression of power on small scales and an enhancement on the large ones. The redistribution of gas ejected from the small scales also contributes to the suppression of power, but it has an antagonist effect on the $P_{\mathrm{Ly}\alpha}$ with the strong energy injection on the large-scale modes. On Fig. 6.8 I show the gas temperature and gas density in HAGN and HnoAGN from the four circled regions from Fig. 6.7, I also show the ratios of the density and temperature of HAGN over HnoAGN. In all cases, we clearly see the hot and dense outflows expeled from galaxies. The outflows extend to larger scales in the HAGN case and the temperature on the edges are ten to hundred times higher when the density of the outflows are about two times higher. Thus, the effect of heating considerably dominates the mass redistribution on the power spectra on the large-scale modes, as seen on Fig. 6.9. The outflows are heated to temperature high enough so that the gas stays in the ionized state, hence reducing the power on the $P_{\mathrm{Ly}\alpha}$. Also the density inside the galaxies are about then time lower in HAGN compared to HnoAGN, confirming the depletion of gas content in galaxies and in their surrounding, hence the reduction of power in the matter power spectra on the small-scale modes.

The redshift dependence is well understood because of the increasing capacity of BH to expel gas from halos on our redshift interval (Beckmann et al., 2017). This is combined with the displacement of energetic gas leading to the expansion of hot-gas bubbles and a net increase of the IGM temperature. The scale dependence arises because of the sensitivity of the power spectrum modes to different regions of the baryonic gas. The large scale modes are sensitive to the diffuse gas, i.e. the efficiently heated region of the temperature-density diagram with temperature above 10^8K, which therefore stays hot and ionized as it is hard for hydrogen to recombine. The small-scale modes are dominated by the signal of dense regions, which are not as efficiently heated as the most diffuse regions, and can partially radiate away the injected energy and ultimately recombine. It therefore alleviates the suppression of power on the smallest scales.

Our results are in agreement with the study done in V13 which shows a suppression of power on the large scales as well. It is also stated that it is due to the heating induced by the AGN feedback for the following reasons. First, the flux PDF exhibits an increase of the number of pixels with high-flux transmittivity, secondly, an increase of low density gas with $T > 10^5$K in the temperature-density diagram is observed. However, corrections from V13 can be 5% larger than the ones derived with HAGN. This is coherent with the fact the OWLS simulations, used in V13, exhibit mass outflow rates one order of magnitude larger than those observed and those in HAGN. Therefore, they overestimate the impact of AGN feedback on the $P_{\mathrm{Ly}\alpha}$ and predict too large suppressions. I recall that uncertainties on the data now reach the percent level precision, suppressing by 5% the $P_{\mathrm{Ly}\alpha}$ is likely to strongly biase the interpretation on cosmological parameters.

6.2.5.2 Uncertainties due to AGN feedback calibration

We identify two possible sources of uncertainties in our correction : uncertainties related to the feedback model or due to the sampling of the LOS.

Uncertainties in the feedback model are related to the uncertainties in the three main subgrid feedback parameters presented in Sec. 6.2.1.1 ; the stochasticity in the accretion rate related to the boost factor α, the efficiency ϵ_{f} and the radius of energy deposition $\mathrm{r}_{\mathrm{AGN}}$. We use the set of additional simulations presented in Sec. 6.2.1.2 to estimate variations in the corrections β at all redshifts due to fluctuations in these parameters. We arbitrary define the 1σ bound due to each parameter variation as

$$\sigma_i = \frac{\beta_i - \beta_{HAGN}}{n}, \tag{6.21}$$

where i is either clp10, clp100, r+, r−, $\epsilon_{\mathrm{f}}+$ or $\epsilon_{\mathrm{f}}-$. We take n as the number of observational uncertainties between the galaxy properties measured in the simulations and the ones from observations. For HAGNclp10 and HAGNclp100 we take $n = 1$ because the deviations between properties of the two simulations and observations are at about the sigma level in terms of observational uncertainties for both the mean fraction of gas and the $\mathrm{M}_{\mathrm{BH}} - \mathrm{M}_*$ relation. For HAGNr+, HAGNr−, HAGNϵ+ and HAGNϵ− we take $n = 3$. Indeed,at least one of the observable is in deviation of at least 3σ in terms of observational uncertainties. We could take $n = 4$ or $n = 5$ if we combined the two probes, but because they are not fully independent and in order to be conservative we choose to keep $n = 3$.

Fig. 6.10 shows the 1σ bound of each of the parameters, at every redshifts :

- **HAGNclp10** and **HAGNclp100** results are presented on the left and right panels of the first row. The stochasticity introduced in the accretion rates of BH appears not to have any noticeable effect on the $P_{\mathrm{Ly}\alpha}$ as the deviations are well below the percent level. Therefore we do not consider any uncertainty due to α in the following.
- **HAGNr+** and **HAGNr−** are on the second row. The parameter $\mathrm{r}_{\mathrm{AGN}}$ comes out to be the one to which $P_{\mathrm{Ly}\alpha}$ is the most sensitive to, with deviations up to 1% for the upper bound and up to 4% for the lower bound. We show that lowering $\mathrm{r}_{\mathrm{AGN}}$ leads to a stronger feedback than increasing it. Indeed, HAGNr− shows a significant decrease

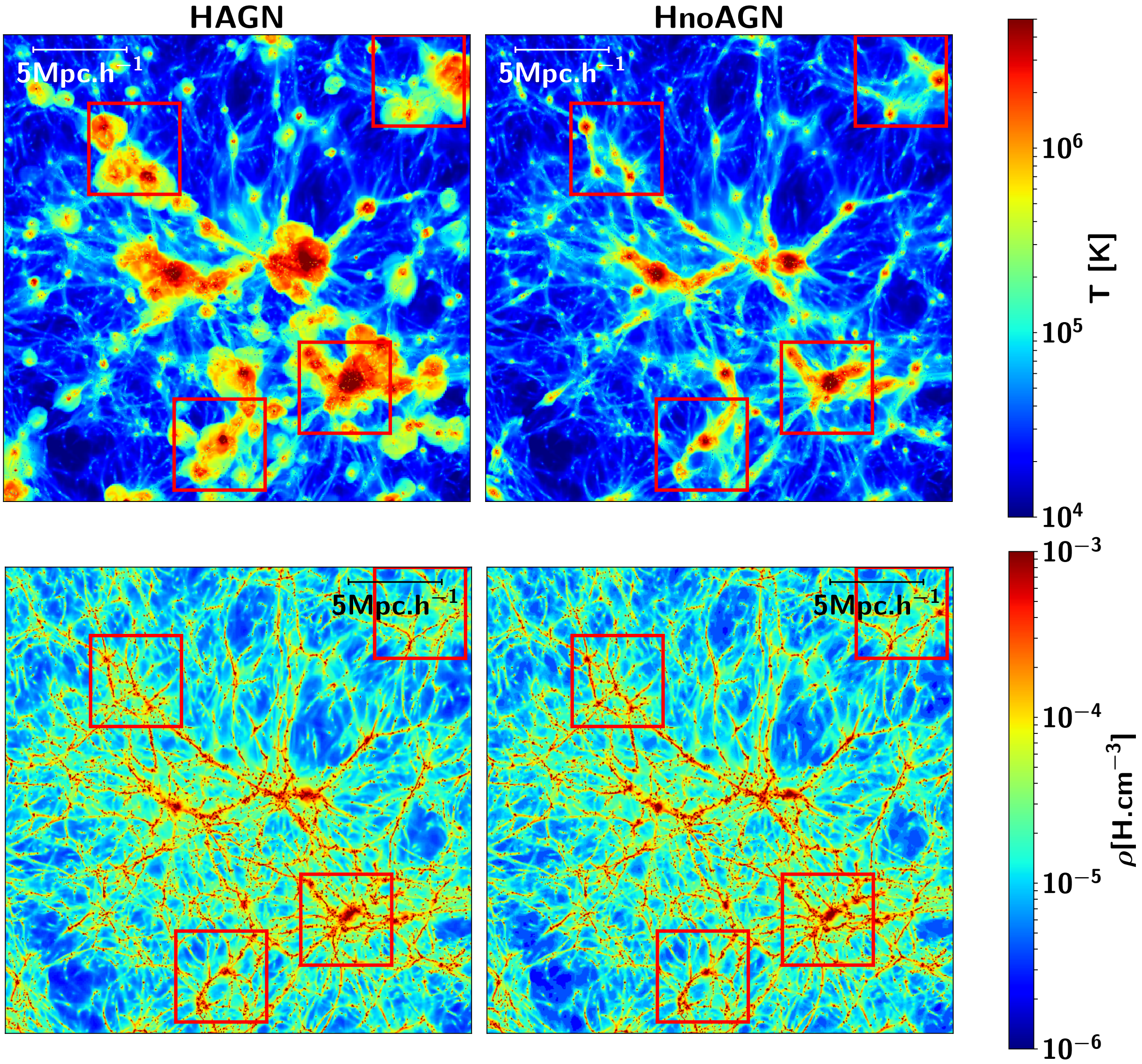

FIGURE 6.7 – *Projected temperature (top) and density (bottom) maps of* HAGN *on the left and* HnoAGN *on the right at* $z = 2.0$ *encoded in* $\log(\mathrm{T})$ *and* $\log(\rho)$ *unit. Boxes are 25* $\mathrm{Mpc}\cdot\mathrm{h}^{-1}$ *in comoving coordinate. The four red boxes are regions A, B, C and D from left to right and top to bottom, I show the zoom of these regions on Fig. 6.8*

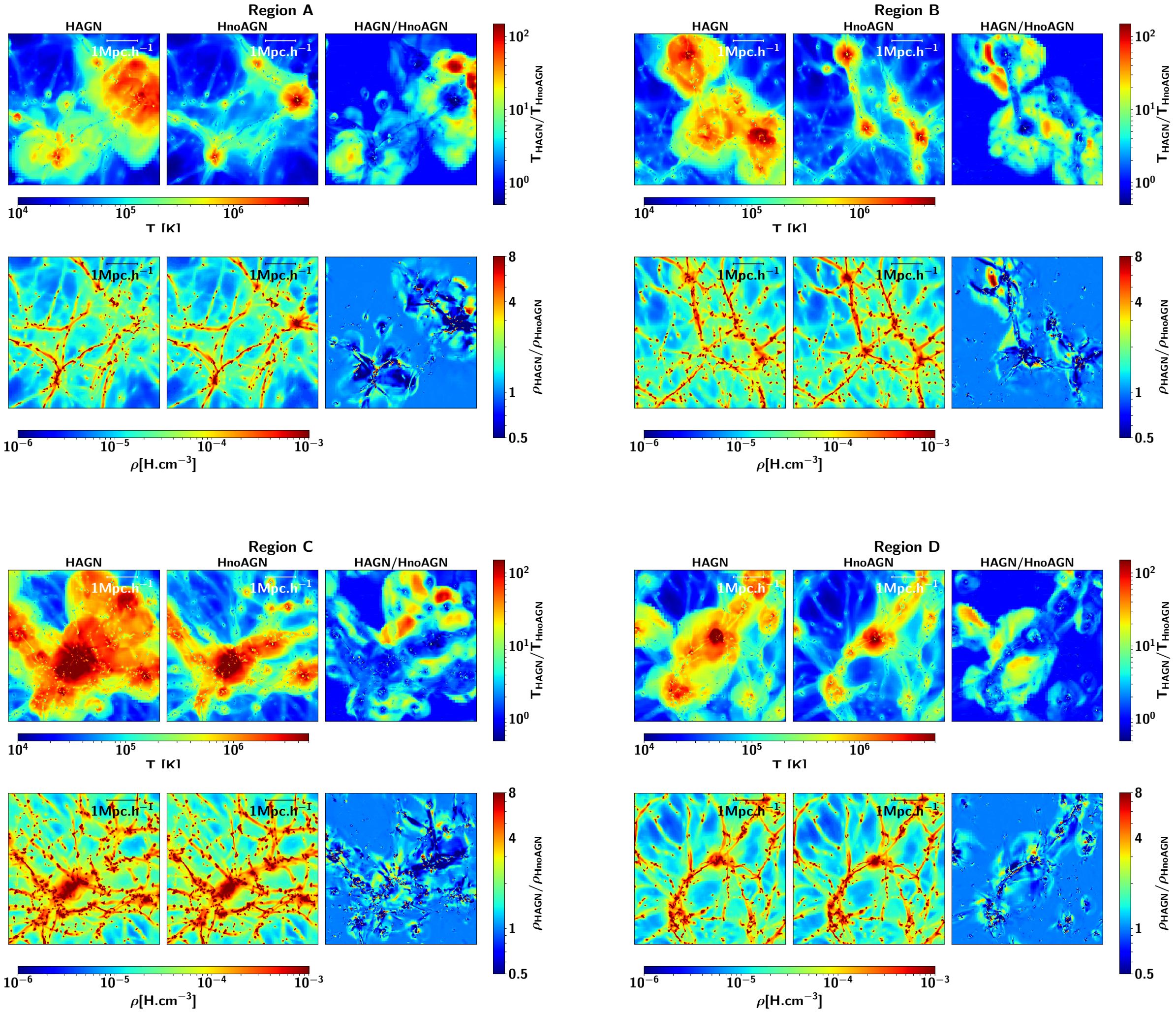

Figure 6.8 – *Projected temperature (first and third lines) and density (second and fourth lines) maps of* HAGN *(first and fourth columns) and* HnoAGN *(second and fifth columns) and the ratio of* HAGN *over* HnoAGN *for the two quantities (third and sixth columns) for the four circled regions from Fig. 6.7. Temperature and density are encoded in* log(T) *and* log(ρ) *unit. Boxes are 5* $\mathrm{Mpc \cdot h^{-1}}$ *in comoving coordinate.*

of power compared to HAGNr+ which displays an increase of power on large scales. Giving more energy to a smaller volume and keeping the same amount of injected energy produces larger hot bubbles of gas around AGNs as we illustrate in the temperature maps of HAGNr+ and HAGNr− at $z = 2$ in Fig. 6.11. There is therefore more ionized gas on large scales when $\mathrm{r_{AGN}}$ is lower. This result is in opposition with Dubois et al. (2012) that shows less ionization for low $\mathrm{r_{AGN}}$. We put this on the account of a different feedback prescription ; in Dubois et al. (2012) the energy injection was volume weighted when it is mass weighted in HAGN. In the first case, when we broaden the region of energy deposition, we impact more diffuse cells that are equally heated than the dense ones and are less likely to radiate away the injected energy, leading to an increase of the ionized region. However in the later case, diffuse cells get less energy than the dense ones, then the dilution makes the feedback less effective.

- **HAGNϵ+** and **HAGNϵ−** are on the third row. Modifications in ϵ_f do not impact the flux power spectrum above the percent level. Even if the $\mathrm{M_{BH} - M_*}$ and $\mathrm{f_{gas}}$ of the two

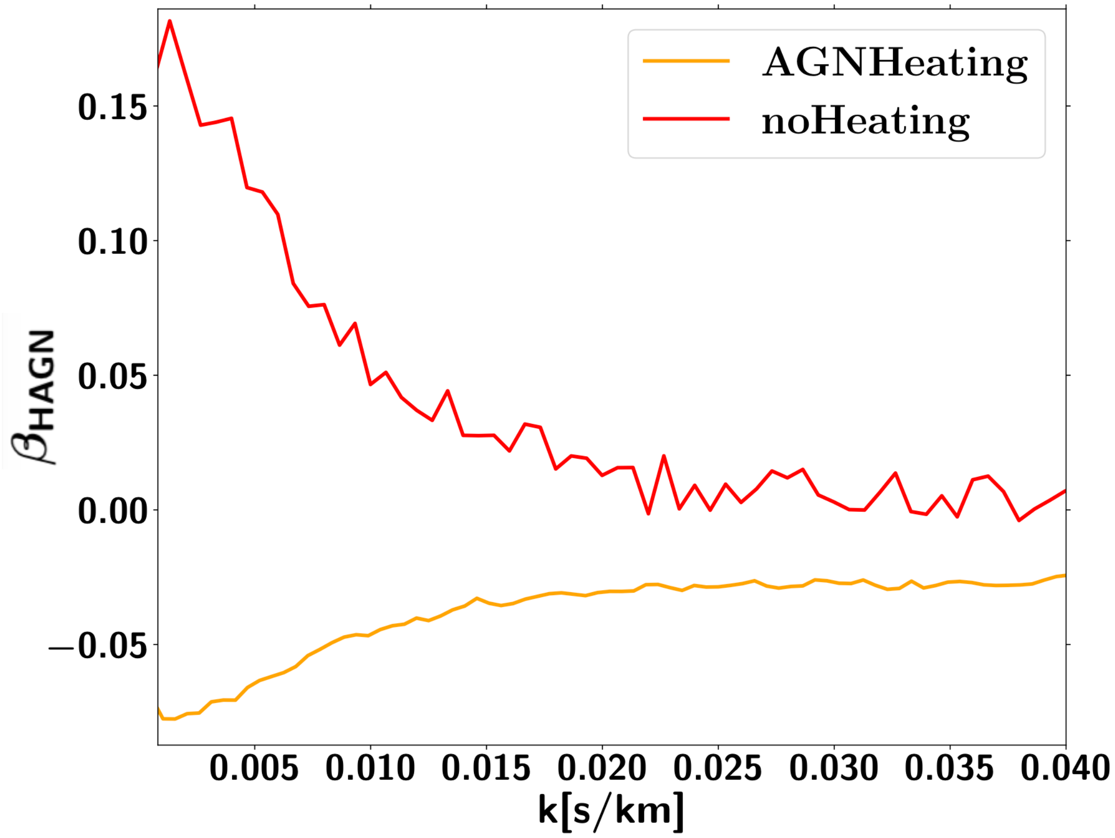

Figure 6.9 – *Correction with and without heating of AGN feedback at $z = 2$. The 'noHeating' curve is obtained by imposing in* HAGN *the same probability distribution function of temperature as a function of of density than in* HnoAGN*, then we compute the neutral fraction of hydrogen.*

resimulations largely differ from observations, the self-regulation of the BH prevents large modifications of the IGM thermal state. We can identify a trend, but since it is largely sub-dominant compared to $\mathrm{r_{AGN}}$ we do not consider any effects due to ϵ_f afterward.

To estimate uncertainties due to the sampling of our LOS sample we compute the root-mean-square (RMS) error of the corrections from five different sets constitued of 20.10^3 LOS. For $z = 4.25$ and $z = 2$ it leads to uncertainties at the level of 10^{-3}. This is subdominant compared to the uncertainties due to $\mathrm{r_{AGN}}$. Therefore we do not consider statistical uncertainties in the following.

We showed that the impact of AGN feedback on the flux power spectra is to globally suppress the power at all scales. The suppression is explained by the combination of an efficient heating and by the mass redistribution from small to large scales. The suppression is enhanced with decreasing redshifts because of the increasing capacity of BH to expel gas from halos and the displacement of hot gas, which induce a stronger feedback. The scale dependence arises because the large-scale modes are sensitive to the diffuse gas and the small-scale modes are dominated by the signal of dense gas that can partly radiate away the injected energy, which alleviates the suppression.

The uncertainty on our correction is strongly dominated by the radius of energy deposition $\mathrm{r_{AGN}}$, because stochasticity in the accretion does not appear to be efficient in our redshift range,

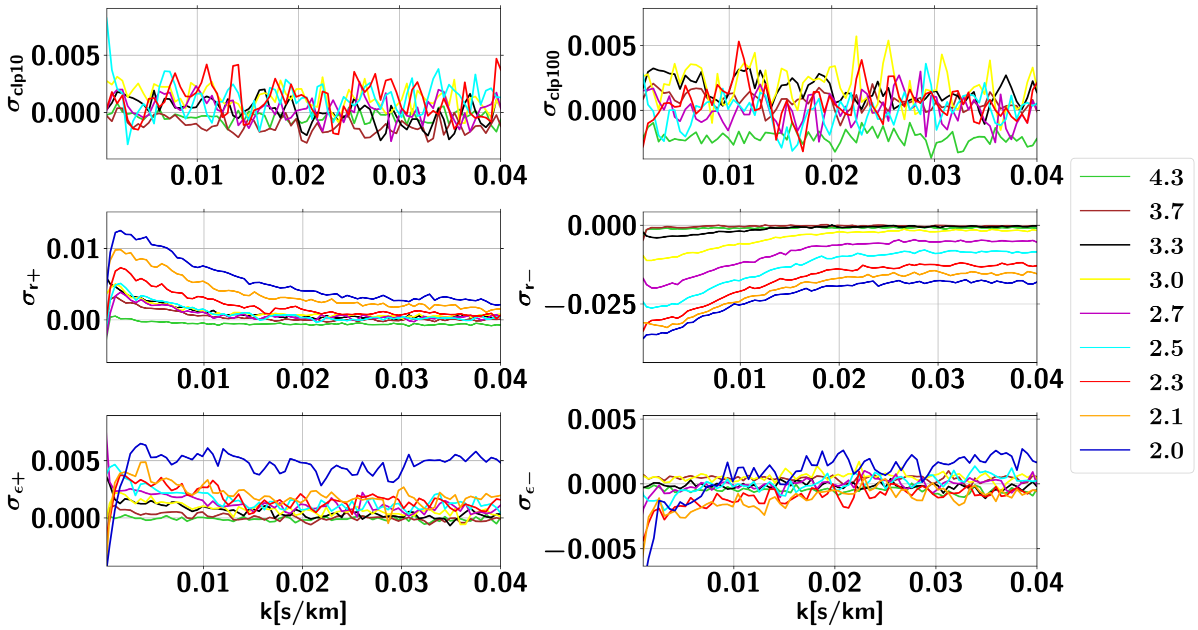

FIGURE 6.10 – *Uncertainties* σ_i *due to the three feedback parameters. From left to right and top to bottom :* HAGNclp10*,* HAGNclp100*,* HAGNr+*,* HAGNr−*,* HAGNϵ+*and* HAGNϵ−*, the different colors are for the nine redshifts.*

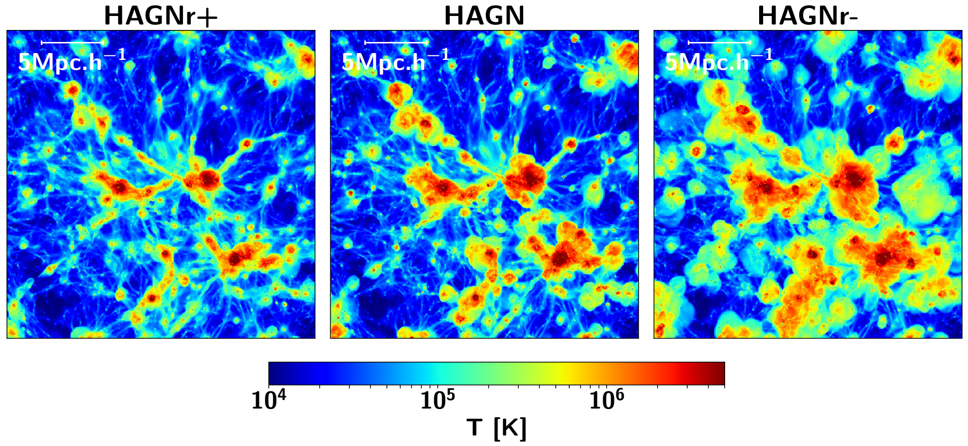

FIGURE 6.11 – *Projected temperature maps of* HAGNr+*,* HAGN *and* HAGNr− *from left to right at* $z = 2.0$ *encoded in* $\log(\mathrm{T})$ *unit. Boxes are 25* $\mathrm{Mpc} \cdot \mathrm{h}^{-1}$ *in comoving coordinate.*

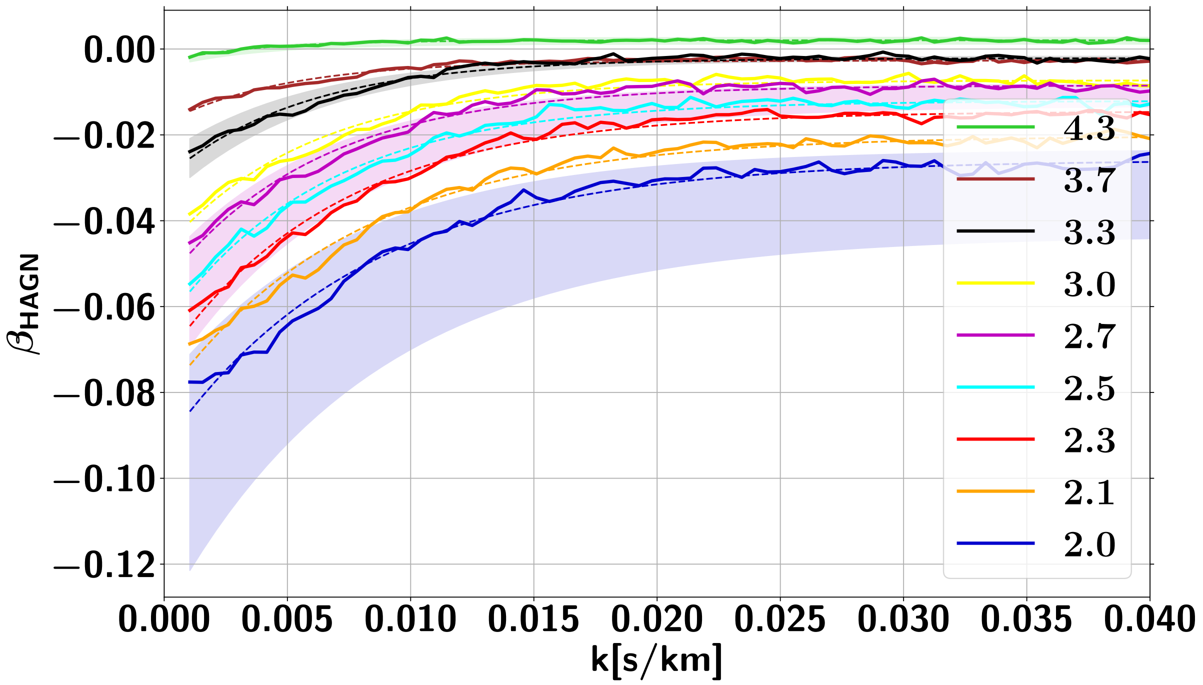

FIGURE 6.12 – *Corrections at all redshits and the associated fits in dashed lines. Uncertainties are displayed in the shaded areas at* $\mathrm{z} = 4.25$*, z3.3,* $\mathrm{z} = 2.7$ *and* $\mathrm{z} = 2.0$ *only for the sake of readibility. Uncertainties are systematics related to uncertainty in the feedback model only.*

and variation of the efficiency is counter-balanced by self-regulation of the BH. It demonstrates that the efficiency of the heating, hence the ionization, has more effect than the amount of injected energy in the medium. I provide analytical fits for the corrections and upper and lower bounds defined as $\beta(\mathrm{HAGN})^{+\sigma_{\mathrm{r}+}}_{-\sigma_{\mathrm{r}-}}$. I fit corrections with the following function, $f(k) = a + b\exp(-ck)$, where the parameters a, b and c are given for the three fits at all redshifts in Tab. 6.4. I show in Fig. 6.12 the fits of the correction in dashed lines, I also display the upper and lower uncertainties in the shaded areas at $\mathrm{z} = 4.25$, $\mathrm{z} = 3.3$, $\mathrm{z} = 2.7$ and $\mathrm{z} = 2.0$ only for the sake of readibility.

6.3 Resolution effects in the IGM : The Extreme-Horizon simulation

Most of the grid-based Lyα studies use uniform grids with 20 $\mathrm{kpc \cdot h^{-1}}$ resolution as it was shown that the IGM Jeans length ($\sim 100\ \mathrm{kpc \cdot h^{-1}}$) should be resolved by at least 4 cells to have the $P_{\mathrm{Ly}\alpha}$ converged (Lukić et al., 2015). However, they do not resolve galaxies, stellar and AGN feedbacks. In order to properly resolve galactic disks and implement feedbacks I used HAGN as the fiducial simulation with an adaptative mesh-strategy : the size of the grid follows the gas density, where the gas is denser the grid cells are smaller (as illustrated in Fig. 5.10). Therefore, HAGN reaches 0.5-1 $\mathrm{kpc \cdot h^{-1}}$ inside galaxies but the resolution is only 100 $\mathrm{kpc \cdot h^{-1}}$ (50 $\mathrm{kpc \cdot h^{-1}}$ respectively) in 80% (20% respectively) of the diffuse IGM. The poor resolution in diffuse regions is common to all state-of-the-art cosmological hydrodynamical simulations implementing AGN feedback, and this is even worse in SPH simulations that do not control resolution in the IGM

Redshift	a	b	c	a_+	b_+	c_+	a_-	b_-	c_-
4.25	$2.03\,10^{-3}$	$-5.03\,10^{-3}$	$2.74\,10^{2}$	$2.71\,10^{-3}$	$-5.23\,10^{-3}$	$2.26\,10^{2}$	$1.16\,10^{-3}$	$-5.49\,10^{-3}$	$2.71\,10^{2}$
3.7	$-2.70\,10^{-3}$	$-1.46\,10^{-2}$	$2.05\,10^{2}$	$-2.60\,10^{-3}$	$-1.07\,10^{-2}$	$2.06\,10^{2}$	$-2.98\,10^{-3}$	$-1.54\,10^{-2}$	$2.30\,10^{2}$
3.3	$-2.13\,10^{-3}$	$-2.81\,10^{-2}$	$1.79\,10^{2}$	$-1.68\,10^{-3}$	$-2.31\,10^{-2}$	$1.77\,10^{2}$	$-2.51\,10^{-3}$	$3.27\,10^{-2}$	$1.70\,10^{2}$
3.0	$-7.29\,10^{-3}$	$-3.94\,10^{-2}$	$1.70\,10^{2}$	$-6.66\,10^{-3}$	$-3.44\,10^{-2}$	$1.64\,10^{2}$	$-8.86\,10^{-3}$	$-5.09\,10^{-2}$	$1.51\,10^{2}$
2.7	$-8.46\,10^{-3}$	$-4.63\,10^{-2}$	$1.61\,10^{2}$	$-8.11\,10^{-3}$	$-4.20\,10^{-2}$	$1.58\,10^{2}$	$-1.32\,10^{-2}$	$-6.47\,10^{-2}$	$1.40\,10^{2}$
2.5	$-1.21\,10^{-2}$	$-5.23\,10^{-2}$	$1.56\,10^{2}$	$-1.19\,10^{-2}$	$-4.65\,10^{-2}$	$1.55\,10^{2}$	$-2.00\,10^{-2}$	$-7.44\,10^{-2}$	$1.34\,10^{2}$
2.3	$-1.45\,10^{-2}$	$-5.81\,10^{-2}$	$1.43\,10^{2}$	$-1.38\,10^{-2}$	$-5.03\,10^{-2}$	$1.45\,10^{2}$	$-2.64\,10^{-2}$	$-8.07\,10^{-2}$	$1.27\,10^{2}$
2.1	$-2.02\,10^{-2}$	$-6.12\,10^{-2}$	$1.30\,10^{2}$	$-1.85\,10^{-2}$	$-5.14\,10^{-2}$	$1.35\,10^{2}$	$-3.48\,10^{-2}$	$-8.32\,10^{-2}$	$1.19\,10^{2}$
2.0	$-2.58\,10^{-2}$	$-6.66\,10^{-2}$	$1.23\,10^{2}$	$-2.35\,10^{-2}$	$-5.48\,10^{-2}$	$1.33\,10^{2}$	$-4.35\,10^{-2}$	$-8.84\,10^{-2}$	$1.20\,10^{2}$

TABLE 6.4 – *Parameters of the correction fits on the function* $f(k) = a + b\exp(-ck)$ *with one line per redshift. The parameters* a, b *and* c *are for corrections given by* HAGN, a_+, b_+ *and* c_+ *are for the upper bound fits, and* a_-, b_- *and* c_- *are for the lower bound fits.*

comoving grid resolution [kpc/h]	97.6	48.8	24.4	12.2	6.1	3.05	1.52	0.76
physical grid resolution [kpc] (z=2)	47	23.5	11.7	5.8	2.9	1.5	0.7	0.3
volume fraction (EH) (z=2)	–	45%	43%	10%	1%	0.04%	$z < 2$	$z < 2$
volume fraction (SH) (z=2)	80%	17%	2%	0.17 %	0.013%	5×10^{-4}%	$z < 2$	$z < 2$
volume fraction (HAGN) (z=2)	77%	19%	2%	0.2 %	0.01%	6×10^{-4}%	$z < 2$	$z < 2$

Table 6.5 – Resolution strategy for EH and SH : the first two lines indicate the comoving and physical (at $z = 2$) grid resolution in kpc/h and kpc, respectively. The last three lines indicate the volume fractions measured at each resolution level at z=2 in EH, SH and HAGN for comparison. In the last two columns $z < 2$ means that these levels are not triggered yet at $z = 2$ but will be for lower redshifts.

(Pillepich et al., 2019). I thus decided to perform convergence tests of the correction through a new simulation, the Extreme-Horizon simulation (EH), which pushes the limits by strongly increasing the resolution in diffuse regions of the IGM and circum-galactic medium (CGM), which I recall, represents 90% of the simulated volume. I co-led the Extreme-Horizon project with collaborators from CEA and IAP, for which we have been allocated about 75 millions CPU hours at the Très Grand Centre de Calcul [3]. I initiated and led the run of the Extreme-Horizon simulation, described in Sec. 6.3.1. I present in Sec. 6.3.2 convergence tests on the analytical AGN feedback correction of the $P_{\mathrm{Ly}\alpha}$. Finally, while my thesis goal was to constrain neutrino masses and to determine the plausibility of a warm dark matter model using the Lyα forest, the EH simulation brings ground breaking results for galaxy-formation mechanisms in cosmological context. I briefly describe results in Sec. 6.3.3, that led to the publication of a Letter in collaboration with all Extreme-Horizon participants, which I included in App. A.

6.3.1 The Extreme-Horizon simulation

The EH simulation is performed with the adaptive mesh refinement code RAMSES (Teyssier, 2002) using the physical models from HAGN (see Sec. 6.2.1.1). The spatial resolution in the CGM and IGM is largely increased compared to HAGN, while the resolution inside galaxies is identical, at the expense of a smaller box size of 50 Mpc $\cdot$ h^{-1}. This justifies the same parameter set of galactic feedback than in HAGN since it only depends on the resolution inside galaxies. The control simulation of the same box with a resolution similar to HAGN is called Standard-Horizon (SH). EH and SH share initial conditions realized with `mpgrafic` [4] (Prunet et al., 2008).

These use a ΛCDM cosmology with matter density $\Omega_{\mathrm{m}} = 0.272$, dark energy density $\Omega_\Lambda = 0.728$, matter power spectrum amplitude $\sigma_8 = 0.81$, baryon density $\Omega_b = 0.0455$, Hubble constant $\mathrm{H}_0 = 70.4$ km$\cdot$ s$^{-1}\cdot$Mpc^{-1}, and scalar spectral index $\mathrm{n_s} = 0.967$, based on the WMAP-7 cosmology (Komatsu et al., 2011). EH was performed on 25,000 cores of the AMD-Rome partition of the Joliot Curie supercomputer at TGCC. Since such a simulation is particularly memory intensive with 3To of data generated at each time step, it was the ideal laboratory to test a new technique of data writing and data reading of massive simulations resulting from a collaboration between the Direction des Applications Militaires (CEA-DAM) and the Departement d'Electroniques des Detecteurs et d'Information pour la Physique (DEDIP, CEA-IRFU) to reduce space-disk usage and to speed-up data access, the `Hercule` parallel I/O library (Bressand et al., 2012; Strafella and Chapon, 2020). It is being run down to $z \sim 0$.

---

[3] http://www-hpc.cea.fr/fr/complexe/tgcc.htm

[4] http://www2.iap.fr/users/pichon/mpgrafic.html

SH uses a 512^3 coarse grid, with a minimal resolution of $100\,\mathrm{kpc\,h^{-1}}$ as in HAGN. Cells are refined up to a resolution of $\simeq 1\,\mathrm{kpc}$ in a quasi-Lagrangian manner : any cell is refined if $\rho_{\mathrm{DM}}\Delta x^3 + (\Omega_b/\Omega_{\mathrm{DM}})\,\rho_{\mathrm{baryon}}\Delta x^3 > \mathrm{m}_{\mathrm{refine,SH}}\mathrm{M}_{\mathrm{DM,res}}$ where ρ_{DM} and ρ_{baryon} are dark matter (DM) and baryon densities respectively in the cell, Δx^3 the cell volume and $\mathrm{m}_{\mathrm{refine,SH}} = 80$. The $\Omega_b/\Omega_{\mathrm{DM}}$ factor ensures that baryons dominate the refinement condition as soon as there is a baryon overdensity. This resolution strategy matches that of HAGN (Table 6.5).

EH uses a 1024^3 coarse grid and a more aggressive refinement strategy with $\mathrm{m}_{\mathrm{refine,EH}} = 1/40\,\mathrm{m}_{\mathrm{refine,SH}}$ in the IGM/CGM (for $\Delta x < 1.52\ \mathrm{kpc}\cdot\mathrm{h}^{-1}$) but similar near and in galaxies : the whole volume is resolved with a twice higher resolution and most of the mass is resolved with a four times higher resolution in 1-D, yielding an improvement of 8 to 64 for the 3-D resolution. This improvement continues until the highest resolution of $\simeq 1\,\mathrm{kpc}$ is reached : the volume fraction at various resolution levels are listed in Table 6.5. Such aggressive approach for grid refinement can better model the early collapse of structures (O'Shea et al., 2005). Fig. 6.13 illustrates the resolution achieved in representative regions of the CGM and IGM in EH and SH. The resolution in EH haloes is typically ∼6 kpc, while it is ∼25 kpc for SH. However, galaxies themselves are treated at the very same resolution in EH and SH : any gas denser than $0.1\,\mathrm{cm}^{-3}$ is resolved at the highest level in SH, as is also the case for 90% of the stellar mass.

Similarly for HAGN, we run companion simulations EHnoAGN and SHnoAGN without AGN feedback. The same seeds are used so we do not have to account for shot noise.

6.3.2 Impact of the IGM resolution on AGN feedback

I construct the 1D power spectrum of the Lyα forest of the same 50,000 LOS in EH and in SH for redshifts $2.0 \leq z \leq 4.2$, following the method described in Sec. 6.2.2.

I show in Fig. 6.15 the ratio of the power spectra between EH and SH rescaled to the same mean flux. The lack of resolution induces a net decrease of power at small scales and an increase of power at large scales up to 15% on the largest scales, which is more significant at high redshifts, similar to what is found in Lukić et al. (2015) (see Fig. 11). Indeed, better resolution at small scales implies to better resolve small-scale collapse, which increases power at small scales. As seen in Sec. 1.3.1.2 while deriving linear equations for the evolution of density perturbations, the Jeans length is the scale at which the gravitational support and pressure force at equal. It needs to be resolved to fully characterize structure collapses. The comoving Jeans length is written as

$$\lambda_J = (1+z)c_s\sqrt{\frac{\pi}{G\rho_m}} = 0.783\sqrt{\frac{T/(10^4\ \mathrm{K})}{\Omega_m(1+\delta)(1+z)}}\ \mathrm{h}^{-1}\cdot\mathrm{Mpc}. \tag{6.22}$$

As such, the redshift evolution of λ_J is determined by the temperature evolution which behaves differently before and after rezionization :

- For $z > z_{\mathrm{reio}}$: the adiabatic expansion sets the temperature to scale as $T \propto (1+z)^2$, thus λ_J decreases with time
- for $z < z_{\mathrm{reio}}$: the temperature is roughly constant with redshift, which is consistent with observations showing that the temperature of the IGM does not change much between $2 < z < 4$ (Becker et al., 2011; Bolton et al., 2014). Thus λ_J increases with time.

Therefore, on the considered range of redshifts here, we see that we need more resolution at high redshifts to resolve Lyα forest absorbers. This explains why resolution effects between EH and SH are larger at high redshift.

Fig. 6.16 shows the differences induced by the net increase of IGM/CGM resolution on the AGN feedback correction, with the ratio of the corrections of the EH simulation, β_{EH}, to the

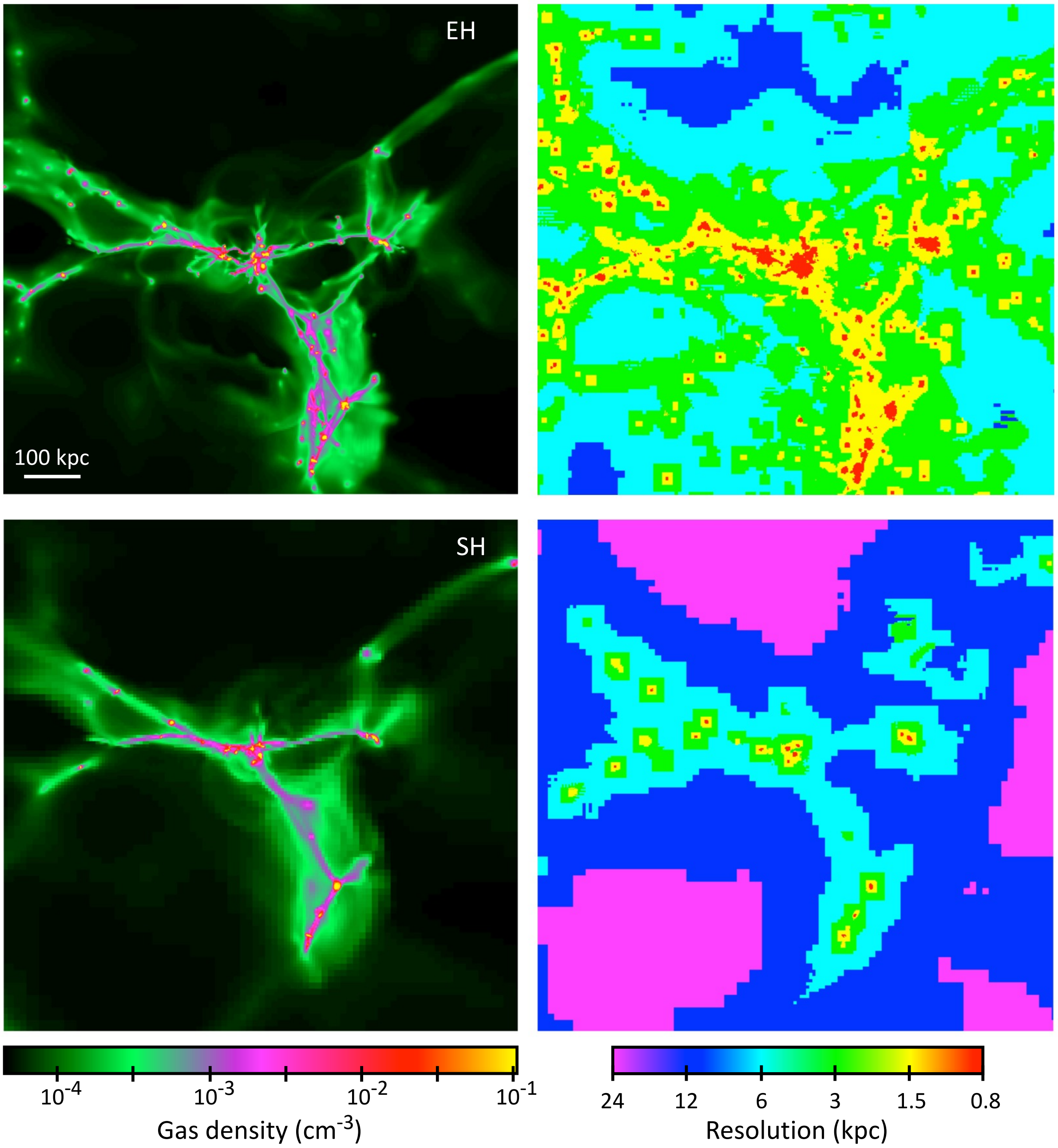

FIGURE 6.13 – *Projected density (left) and physical resolution (right) in EH (top) and SH (bottom) zoomed on a massive halo at* $z = 3$*. The depth of the projections are 200* $\mathrm{kpc} \cdot \mathrm{h}^{-1}$ *and the boxes extend 1* $\mathrm{Mpc} \cdot \mathrm{h}^{-1}$ *on each side. The gas density is computed as the mass-weighted average of local densities along the line-of-sight corresponding to each pixel. The resolution shown is the resolution of the cell in which the gas density is the highest along each line-of-sight.*

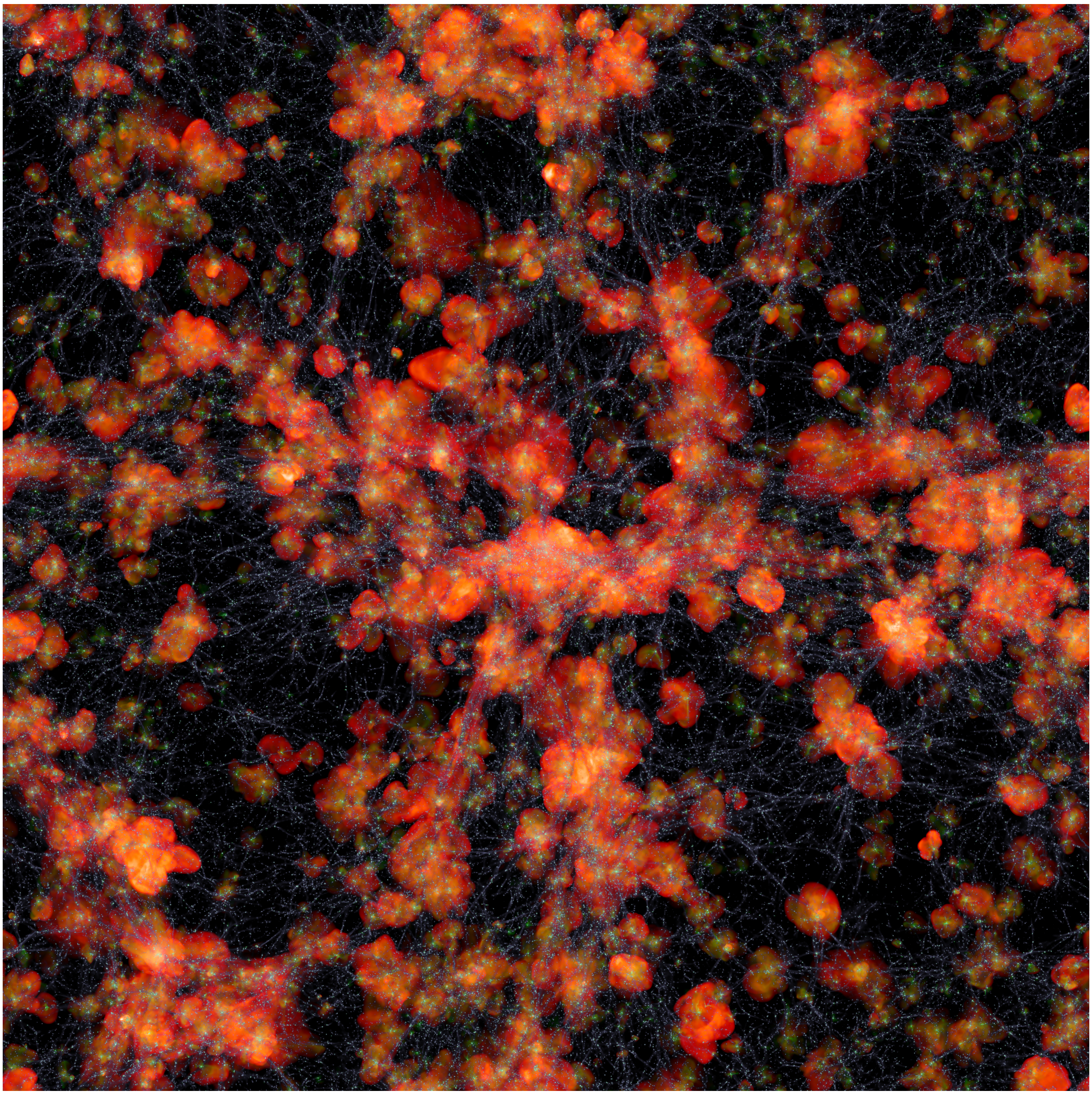

FIGURE 6.14 – *Projected map of the EH simulation at* $z \simeq 2$. *Gas density (grey), entropy (red) and metallicity (green) are shown.*

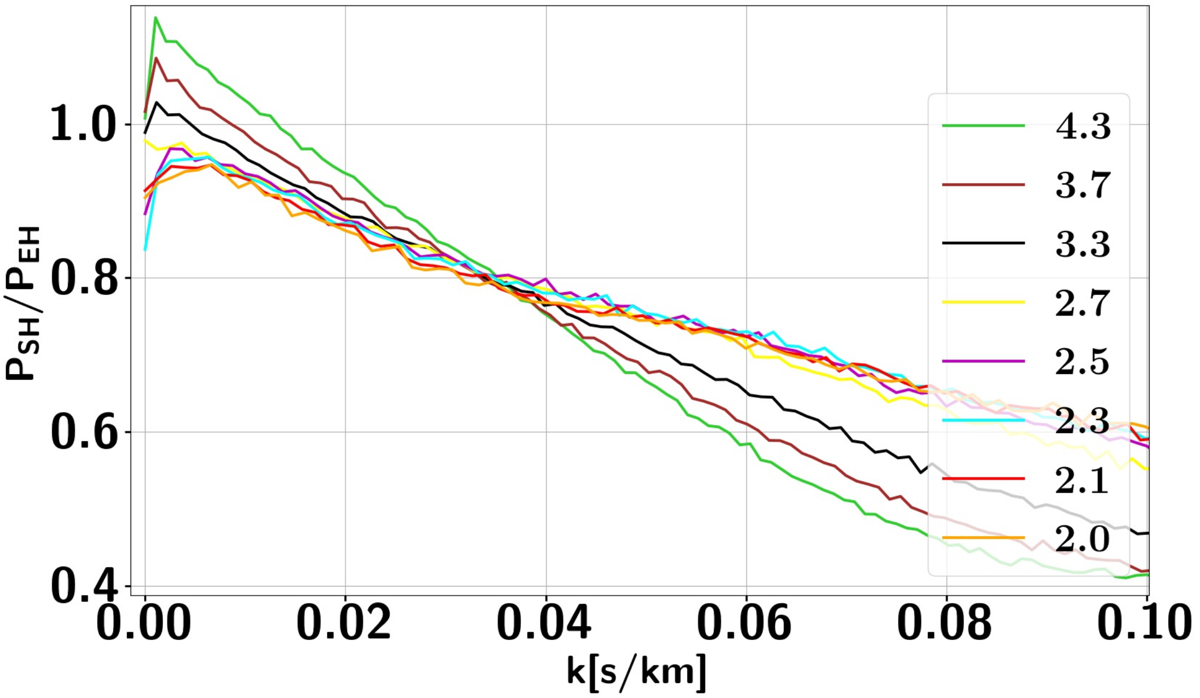

FIGURE 6.15 – *Ratio of the* $P_{\mathrm{Ly}\alpha}$ *from SH and EH at different redshifts on a wide range of scales. BOSS and eBOSS go down to 0.02* $\mathrm{km} \cdot \mathrm{s}^{-1}$ *while DESI reaches 0.04* $\mathrm{km} \cdot \mathrm{s}^{-1}$.

fiducial correction β_{HAGN}. The resolution effect is well below the percent level at every redshifts. We stress that it does not imply that the absolute $P_{\mathrm{Ly}\alpha}$ are converged, as seen in Fig. 6.15, but it means that the coupling of AGN feedback and resolution is greatly subdominant when compared to other sources of uncertainties. We can therefore consider afterward that the AGN feedback corrections estimated with HAGN are converged on our range of scales and redshifts.

6.3.3 Impact on galaxy-formation mechanisms

While my main interest in producing EH was to probe resolution effects in the IGM and CGM regarding the Lyα forest, it brings ground-breaking results for galaxy-formation mechanisms in comsmological contexts.

First, the high resolution in low-density regions results in smaller-size massive galaxies at redshift $z = 2$, in better agreement with observations compared to other simulations. I will briefly describe the results of this finding below, which led to the publication of a Letter with all the Extreme-Horizon participants included in App. A.

In left panel of Fig. 6.17, I show stellar masses and radii of massive galaxies from EH and SH at z$\sim$ 2 compared to the model from Dutton et al. (2011), known to provide a good fit to Main Sequence (MS) galaxies. SH galaxies are larger than both EH galaxies and observed MS galaxies. EH galaxies generally lie around the observed relation, and a small fraction have significantly smaller sizes. We define the compactness $\mathcal{C}$ as the ratio between the radius expected from the Dutton et al. (2011) model and the actual radius. The compactness distribution for EH (right panel of Fig. 6.17) peaks at around $\mathcal{C} \simeq 1$ but exhibits a distinct tail for $\mathcal{C} > 1.3$. We thus define two massive galaxy populations in EH : 10 ultra-compact (UC) galaxies with $\mathcal{C} > 1.3$ and 50 non ultra-compact (NUC) ones.

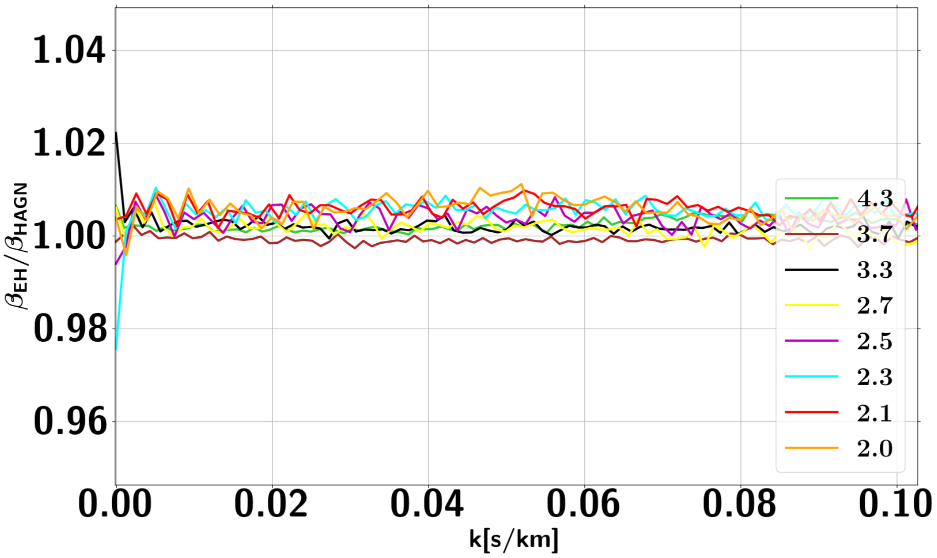

FIGURE 6.16 – *Ratio of the AGN feedback corrections on the* $P_{\mathrm{Ly}\alpha}$*, using HAGN and the more IGM-resolved simulation, EH. This demonstrates the convergence of the previously established analytical correction on a wide range of scales, even smaller than DESI will reach.*

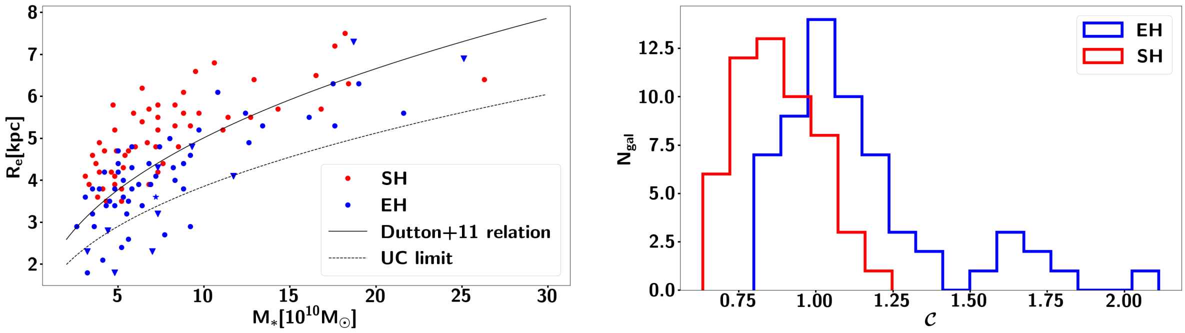

FIGURE 6.17 – ***Left :*** *Stellar half-mass radius* R_e *versus stellar mass* M_* *for massive galaxies at* $z = 2$ *in EH and SH. The displayed model from Dutton et al. (2011) provides a good fit to Star Forming Galaxies (SFG) at z=2. UC galaxies lie below the black dashed line while NUC galaxies are above. We identify EH galaxies above and below the Main Sequence of star formation (MS) with stars and triangles, respectively, following the definition of the MS from Schreiber et al. (2017).* ***Right :*** *Compactness distributions for the EH and SH massive galaxies at* $z = 2$.

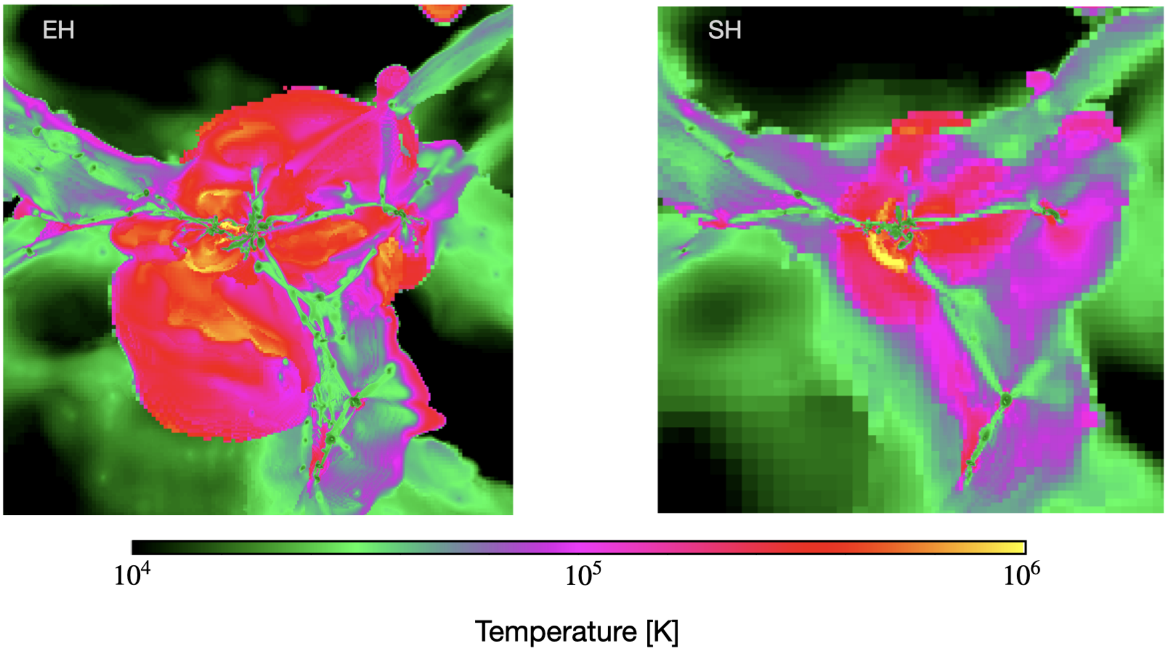

FIGURE 6.18 – *Projected temperature in EH (left) and SH (right) zoomed on a massive halo at* $z = 3$*. The depth of the projections are 200* $\mathrm{kpc \cdot h^{-1}}$ *and the boxes extend 1* $\mathrm{Mpc \cdot h^{-1}}$ *on each side. The temperature shown is the temperature of the cell in which the gas density is the highest along each line-of-sight. For better lisibility, the minimal temperature shown is set to* 10^4 *K and the maximal temperature shown is* 10^6 *K.*

Hence, massive galaxies in EH are globally more compact than in SH, and EH contains a population of UC outliers. The size difference is not expected to arise from internal processes such as instabilities and/or feedback, as galactic scales and feedback are treated with the very same resolution in EH and SH. On the first hand, we show the importance of modeling diffuse gas flows at high-enough resolution in the IGM and CGM, as this tends to reduce the angular momentum supply onto massive galaxies. On the other hand, the formation of UC galaxies does not result from diffuse gas accretion but from repeated major mergers of low-mass progenitors. A pleasant outcome of our analysis is that issues in galaxy formation simulations could indeed be solved by accurately resolving structure formation without calling upon feedback or novel subgrid models.

Finally, hot halo gas around massive galaxies are more extended when the IGM is better resolved in EH compared to SH, as I show in Fig. 6.18. Stronger feedbacks seem not plausible, as for the mechanisms explaining more compact galaxies. Indeed, baryonic feedbacks are treated at the very same resolution. Also, supermassive black holes in EH are below 10% more massive than in SH. Therefore AGN feedback do not inject more than 10% additional energy in EH compared to SH injected if we take into account the Eddington limit, while differences in temperature reaches 200% to 300%. Possible explanations could come from a better resolution of heating/cooling processes and/or a better resolution of the spatial position of virial shocks. This tentatively impacts gas accretion onto galaxies and the very properties of galaxies themselves. Indeed, hot halos can induce galaxy quenching (i.e. the galaxy is turining from blue and star-forming to red and passive).

7

Cosmological Results

Contents

THE power spectrum of the Lyα forest stands as an invaluable probe for constraining cosmology. It is not just an additional tool to confirm already existing results ; because it probes scales significantly smaller than other cosmological probes, and because it opens a window on a range of redshifts where the evolution of fluctuations are not erased by non linearities yet, the Lyα forest has the power to make important steps toward a more precise understanding of our universe. First, by probing small scales, it is sensitive to the smoothing of fluctuations induced by the free streaming of relativistic particles, such as active neutrinos or thermal warm dark matter. In particular, the information embedded in these smallest scales enable to break degeneracies when combining with other cosmological probes that only the Lyα forest can break, and therefore significantly thighten constraints. Also, Lyα analysis can reveal tensions when confronted with different probes, which potentially brings to light the need of new physics beyond the standard model. By analogy with the H_0 tension between late and early universe, the Lyα forest appears to reveal tensions on n_s between the late and early universe, also confirmed by weak lensing analysis.

So far, on the observational side, I presented the measurements of the 1D power spectrum of the Lyα forest using SDSS data, along with a thorough study of the associated systematics. On the theoretical side, I explained the necessity of hydrodynamical simulations to get robust theoretical predictions. I described the existing grids of simulations covering the cosmological parameter space of neutrinos and thermal dark matter. In particular, I presented the analysis of the impact of AGN feedback on the Lyα forest to significantly improve these theoretical predictions and to reach a similar level of precision than in the data.

We now have all the tools, observations and theoretical predictions, which we can combine to constrain cosmology. I will proceed in two classical ways. In Sec. 7.1, I will invert the signal from many cosmological observables, along with the Lyα forest signal, to estimate the total matter power spectrum at $z = 0$ and provide a qualitative consistency test of the ΛCDM model. In section 7.2 I will directly use the Lyα signal to constrain cosmological parameters, in particular the sum of the masses of active neutrinos and of thermal relic warm dark matter and discuss a mild tension between Lyα and CMB data.

7.1 The matter power spectrum : from Lyα forest to CMB scales

The ΛCDM model provides a simple and remarkable fit to much of the existing cosmological data, forming the basis of the standard cosmological paradigm. The CMB temperature and polarization anisotropies observed by the *Planck* satellite can be explained with only the six free parameters of the ΛCDM model (Planck Collaboration et al., 2018c; Collaboration et al., 2018). In this section, I illustrate the extent to which this model, with parameters fixed to their best-fit given *Planck* data, is in agreement with a number of other probes spanning cosmic time and cosmic scales.

In an initial work, Tegmark and Zaldarriaga (2002) demonstrated the consistency between the ΛCDM model fit to the WMAP CMB data (Bennett et al., 2013), the first iteration of the SDSS (SDSS I) (York et al., 2000) clustering data that were available at the time, the 2 Degree Field Galaxy redshift Survey(2dFGRS) (Colless et al., 2001) galaxy clustering data and the Red-Sequence Cluster Survey (Hoekstra et al., 2002) weak lensing data. More recent updates to this work include Tegmark and Zaldarriaga (2009) and Hlozek et al. (2012), which included newer data and other types of probes. With the advent of the Planck mission, of the third and fourth iterations of the Sloan Digital Sky Survey (Blanton et al., 2017) and of the Dark Energy Survey (The Dark Energy Survey Collaboration, 2005), the measurements have now reached an improvement of about an order of magnitude in precision over the last two decades since the initial work. These updated data sets make it timely to reevaluate the overall agreement. The analysis presented below led to the publication Chabanier et al. (2019) in collaboration with my PhD advisor Nathalie Palanque-Delabrouille and Marius Millea from the Planck collaboration.

The main results of the analysis are two-fold. First, focusing in particular on the Lyα constraints, I developped a new more accurate method for processing these data into a constraint on the linear matter power spectrum, $P_{\mathrm{m}}(k)$, at redshift zero. This method is based on a technique known as total variation regularization (TVR ; Chartrand, 2005), which reduces noise in the resulting estimate. Second, we take this constraint, combined with a number of others, and produce a compilation of $P_{\mathrm{m}}(k)$, shown in Fig. 7.3. On scales of a few Mpc, we include the information embedded in the Lyα forest measured with the quasar survey of the SDSS IV DR14 (Abolfathi et al., 2018). Partially overlapping in scale, we also use the cosmic shear measurement from the DES YR1 data release (Troxel et al., 2017). On scales of several tens of Mpc, we use the power spectrum of the halo density field derived from a sample of luminous red galaxies (LRG) from the SDSS DR7 (Reid et al., 2010). Finally, on the largest scales, we use the anisotropies of the CMB measured by the Planck satellite. In addition to probing a wide range of scales, from $k = 2 \times 10^{-4}$ to $k = 2\ \mathrm{h} \cdot \mathrm{Mpc}^{-1}$, these data also cover a large range of cosmic epochs : $z \sim 0.35$ for the LRG, $z \sim 0.2$ to 1.3 for the shear measurements, $z = 2.2$ to 4.6 for the more distant Lyα forest, and $z \sim 10^3$ for CMB.

As described in Tegmark and Zaldarriaga (2002), inferring the linear matter spectrum at $z = 0$ from the various probes we consider here is a highly model-dependent process. We take as our fiducial model the Planck 2018 best-fit ΛCDM model (Planck Collaboration et al., 2018c). The results here are therefore a test of the consistency of this model, rather than direct constraints on the matter power spectrum. In general, we find qualitative agreement of this fiducial model with the data we consider.

The datasets which we consider were chosen to be representative of different types of cosmological measurements which exist and to cover a broad range of scales, particularly favoring ones where data products were especially convenient for the calculations we perform here. Of course, many other measurements exist which provide constraints on the matter power spectrum, some of which are known to be in varying degrees of tension with the *Planck* best-fit model. It is beyond the scope of this work to include them all, however we provide a Dockerized Jupyter notebook which includes the fairly complex dependencies needed to produce this plot. We hope that this makes it easy for any group in the future to add any desired data set and keep up-to-date this compilation. The repository for this notebook can be found here : [1].

Sec. 7.1.1 presents the Lyα data and explain how I compute the 3D matter power spectrum from measurements derived in Chapter 4. These data are the ones whose treatment differs the most from the previous study of Croft et al. (2002) used in Tegmark and Zaldarriaga (2002). Sec. 7.1.2 presents the other probes we use (CMB, cosmic shear and galaxy clustering) and the general method we apply to compute the 3D matter power spectrum in each case. General interpretation is outlined in Sec. 7.1.3.

---

[1] `https://github.com/marius311/mpk_compilation`

7.1.1 From the Lyα forest

7.1.1.1 Lyman-alpha data

I use the 1D power spectra of the Lyα forest from the DR14 of the BOSS/eBOSS programs of the SDSS derived in Chapter 4. The analysis gives an estimate the 1D Lyα power spectrum along a line of sight, $P_{\mathrm{Ly}\alpha}$, in thirteen equally-spaced redshift bins covering the range $z = 2.2$ to 4.6 with $\Delta z = 0.2$. The highest redshift bin is built from 63 quasars only and has large uncertainties. I therefore use only the lowest twelve redshift bins here.

I recall that these data show an oscillatory feature due to the correlated absorption by Lyα and Si III at a velocity separation $\Delta v = 2271\ \mathrm{km}\cdot\mathrm{s}^{-1}$. Indeed, our estimator P_{1D} contains the signal of both absorptions that could not be separated at the data analysis step. Adopting the approach from McDonald et al. (2006), we model the transmission flux fraction of Lyα absorptions and Si III absorptions as,

$$\delta_{\mathrm{Ly}\alpha+\mathrm{Si}} = \delta_{\mathrm{Ly}\alpha}(v) + a\delta_{\mathrm{Ly}\alpha}(v+\Delta v), \tag{7.1}$$

with $\delta_{\mathrm{Ly}\alpha}(v)$ being only for Lyα. The resulting power spectrum is

$$P_{\mathrm{1D}}(k) = (1+a^2)P_{\mathrm{Ly}\alpha}(k) + 2a\cos(\Delta v\,k)P_{\mathrm{Ly}\alpha}(k) \tag{7.2}$$

We use equation (7.2) to correct for these wiggles, where a is fit independently for each redshift bin. We use these 1D Lyα power spectra to derive the 3D matter power spectrum as explained below.

7.1.1.2 Method

I follow the prescription of Croft et al. (1998), updated in Croft et al. (2002). We assume that the 3D Lyα power spectrum $P_{\mathrm{Ly}\alpha,\mathrm{3D}}$ is related to the linear matter power spectrum P_{m} by a proportionality relation,

$$P_{\mathrm{m}}(k,z) = \frac{P_{\mathrm{Ly}\alpha,\mathrm{3D}}(k,z)}{b^2(k,z)}, \tag{7.3}$$

with $b(k,z)$ a scale and redshift dependent bias that depends on the cosmological model. The scale dependence is an improvement over the initial methodology, added in Croft et al. (2002), to take into account the effects of non-linear evolution, thermal broadening and peculiar velocities.

The 1D and the 3D Lyα power spectra are related by

$$P_{\mathrm{Ly}\alpha,\mathrm{3D}}(k) = -\frac{2\pi}{k}\frac{\mathrm{d}P_{\mathrm{Ly}\alpha}(k)}{\mathrm{d}k}\ , \tag{7.4}$$

which we use to derive the 3D Lyα power spectrum needed in equation (7.3).

I compute the bias $b(k,z)$ for each of the twelve redshift bins mentioned above using CAMB [2] (Lewis et al., 2000) for the linear matter power spectrum. For the 1D Lyα power spectrum, I use hydrodynamical simulations from *G_BASE* grid covering the flat ΛCDM parameter space described in Sec. 6.1 of Chapter 6.

The dependence of the power spectrum with cosmological and astrophysical parameters is modeled by a Taylor expansion of hydrodynamical simulation power spectra. The grid of simulations required for this interpolation consists of a reference simulation, called the best-guess simulation, centered on the Planck 2013 best-fit cosmology Ade et al. (2014) and an IGM

---

[2] `https://camb.info`

Table 7.1 – *Fit parameters. First column : central value and variation range in the simulation grid. Second column : best-fit value and 68% confidence interval for a fit to Lyα + Planck (TT + lowE).*

Parameter	Simulations	Best-fit
n_s	0.96 ± 0.05	0.954 ± 0.004
σ_8	0.83 ± 0.05	0.817 ± 0.007
Ω_m	0.31 ± 0.05	0.330 ± 0.009
H_0 $(km.s^{-1}.Mpc^{-1})$	67.5 ± 5	66.2 ± 0.6
$T_0(z=3)$ (K)	14000 ± 7000	11300 ± 1600
$\gamma(z=3)$	1.3 ± 0.3	0.7 ± 0.1
A^τ	0.0025 ± 0.0020	0.0026 ± 0.0001
η^τ	3.7 ± 0.4	3.734 ± 0.015

thermal history in agreement with Meiksin (2009); Becker et al. (2011), completed by simulations where one or two parameters at a time are given off-centered values. These simulations are used to compute a full second-order Taylor expansion around the Lyα flux power spectrum measured for the reference case to test other cosmologies.

In Tab. 7.1, we list some of the values of the parameters used in the best-guess simulation of B14, as well as the corresponding best-fit values measured in Chabanier et al. (2019), for a fit to the eBOSS 1D Lyα power spectrum combined with the Planck 2018 “TT+lowE" likelihood (Planck Collaboration et al., 2018c). The cosmological analysis to perform this fit will be fully described in Sec. 7.2. The best-fit model is in good agreement with the central best-guess simulation. The parameters that deviate the most from their central value are σ_8 and Ω_m. We determine the biases b_{bf} for the best-fit model by computing the biases b_{bg} for the best-guess simulation, and we apply first-order corrections to account for the measured shifts in σ_8 and Ω_m, using simulations where all parameters are kept to their central value except for either σ_8 or Ω_m. We determine the bias $b(z,k)$ at each redshift z and scale k by

$$\begin{aligned} b_{\mathrm{bf}}(z,k) = b_{\mathrm{bg}}(z,k) & \\ &+ (\sigma_{8,\mathrm{bf}} - \sigma_{8,\mathrm{bg}}) \frac{\mathrm{d}b}{\mathrm{d}\sigma_8}(\sigma_{8,\mathrm{bg}}, \Omega_{m,\mathrm{bg}}) \\ &+ (\Omega_{m,\mathrm{bf}} - \Omega_{m,\mathrm{bg}}) \frac{\mathrm{d}b}{\mathrm{d}\Omega_m}(\sigma_{8,\mathrm{bg}}, \Omega_{m,\mathrm{bg}})\ . \end{aligned}$$

Left panel of Fig. 7.1 shows both best-guess and best-fit biases for redshift $z = 2.8$. As illustrated in the figure for a specific redshift, but similarly for all redshifts, the linear corrections have little effect. Equation (7.3) thus allows us compute the linear power spectra $P_{\mathrm{m}}(k, z_i)$ for all twelve redshift bins z_i. We then transpose each of them to $z = 0$ with the relation

$$P_{\mathrm{m},z_i}(k,0) = P_{\mathrm{m}}(k, z_i) \times t(k, z_i)\ , \tag{7.5}$$

where the evolution term $t(k, z_i)$ is determined in linear theory using a Boltzmann code such as CAMB [3] (Lewis et al., 2000) or CLASS [4] (Lesgourgues, 2011). Finally we combine all twelve $z = 0$ power spectra P_{m,z_i} using an inverse-variance weighted average. Top right panel of Fig. 7.1 shows the resulting $P_{\mathrm{m}}(k, 0)$.

---

[3] `https://camb.info`

[4] `http://class-code.net/`

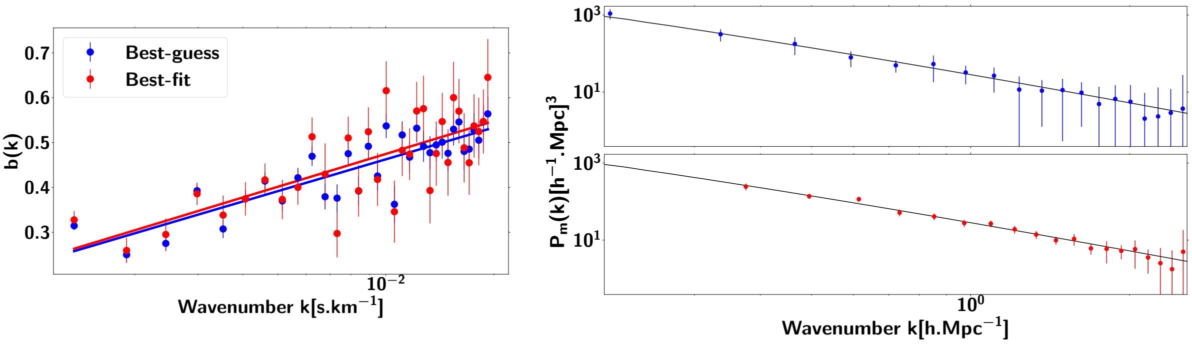

FIGURE 7.1 – ***Left :*** *Biases computed at* $z = 2.8$ *for the best-guess (in blue) and best-fit (in red) configurations. The lines are linear-log fits to each case.* ***Right :*** *Linear matter power spectrum inferred from Ly*α *data. Results from the discrete differentiation are shown in the top panel, and from the TVR approach in the bottom panel. The black solid line is the linear theory expectation.*

I stress that we do not use any analytical corrections for baryonic feedback since these are second order corrections in the computation of the biases. These are fully encompassed by the large uncertainties.

7.1.1.3 Total Variation Regularization

The discrete differentiation of the 1D Lyα power spectrum $P_{\mathrm{Ly}\alpha}$ to obtain the 3D transmitted flux power spectrum $P_{\mathrm{Ly}\alpha,\mathrm{3D}}$ significantly amplifies noise and uncertainties. The effect is worst at small scales where only the highest redshift bins, which are also the noisiest, contribute to the measurement. To reduce this computational artifact, we use a refined differentiating technique, the total variation regularization (TVR) method, proposed in Chartrand (2005). It is a specific regularization process that estimates the derivative of a function f as the minimizer u_{min} of the functional F,

$$F(u) = \alpha R(u) + DF(Au - f), \tag{7.6}$$

where α is the regularization parameter, $R(u)$ is the regularization term which penalizes noise, and $DF(Au - f)$ is the data fidelity term with $Au(x) = \int_0^x u$. The TVR uses $R(u) = \int |u'|$ and $DF(\cdot) = \int |\cdot|^2$. The resulting algorithm has only one free parameter, α, that we fix to 10^{-5} for all the redshift bins, as it appears to be a good compromise between smoothing the data and conserving valuable information. We tested the TVR on an analytical form of the 1D Lyα power spectrum, which allowed us to compare the resulting derivative to the true $P_{\mathrm{Ly}\alpha,\mathrm{3D}}$. The TVR induces no computational bias, except on the first three sampling points, which we hence decide not to keep in the following. To estimate the uncertainty on the 3D power spectrum resulting from this regularization, I perform a parametric bootstrap at each k bin with 1000 iterations. The right bottom panel of Fig. 7.1 shows the final 3D matter power spectrum at redshift $z = 0$ derived with the TVR approach. The dispersion is clearly reduced and the power spectrum from TVR considerably smoother than the one from a straight derivative. The TVR technique increases the correlations between neighboring points (up to 50% in the worst case, for nearest-neighbor correlation), although correlations with next-to-nearest neighbors are between 1 and 20% at most.

Finally, I point out that we use the TVR derivation for the data but we keep to straight derivatives to compute the biasing functions from the hydrodynamic simulations. The reason is the following. The $P_{\mathrm{Ly}\alpha}$ from the simulations is much smoother than in the data, and systematic uncertainties from the bias term are largely sub-dominant compared to data statistical uncertainties. Using the TVR technique on the simulations would therefore unnecessarily increase the correlations between neighboring points without yielding a measurable gain on the resulting uncertainties.

7.1.2 From other probes

Having described in some detail the Lyα forest constraints and our new TVR-based method for calculating them, we now turn to constraints from the other datasets considered, which more closely follow the procedure laid out by Tegmark and Zaldarriaga (2002). Their procedure is based on the relating a given observable, d_i (which can be for example a CMB C_ℓ, or measurement of cosmic shear power spectrum at some redshift, etc...), to $P_{\rm m}(k,0)$, via

$$d_i = \int d\ln k\, W_i(k) P_{\rm m}(k,0) \tag{7.7}$$

Each given observable will have a different window function, $W_i(k)$, which can be calculated from theory for a fixed cosmological model. In many cases, for example if our d_i are simple auto-correlation functions, the $W_i(k)$ are strictly positive. Furthermore, depending on the exact quantity measured, they are often also fairly localized in k. In these cases, we normalize the $W_i(k)$ to unit area, effectively treating it as a probability distribution, and, following Tegmark and Zaldarriaga (2002), take the error bar in the k-direction in Fig. 7.3 to denote the middle 80% quantile of this distribution. Our slight modification to their procedure is that whereas they take the middle 80% of the quantity $W_i(k)P_{\rm m}(k,0)$, we take it of just $W_i(k)$. We view this as the more natural choice since it is just $W_i(k)$ which represents the projection of the data into the redshift zero matter power spectrum. Additionally, this gives us a k-direction error bar which does not depend on the shape of $P_{\rm m}(k,0)$.

In Fig. 7.2, we plot the window functions for the different observations which we use. In each case, some "rebinning" of the data is applied as compared to the raw data products provided by each experiment. This is done so as to produce more reasonably spaced data points in the k direction, and to improve the localization of the $W_i(k)$. We describe these rebinnings in the individual sections below. One can verify the localized nature of the different window functions, indicating the validity of interpreting each data point as a constraint on $P_{\rm m}(k,0)$.

Cosmic microwave background

For CMB data, we use the *Planck*2018 temperature, polarization, and lensing reconstruction power spectra (Collaboration et al., 2018,?). At $\ell < 30$ in temperature, we use the C_ℓ's provided by the COMMANDER likelihood, with the asymmetric errorbars averaged together, which should have minimal impact as we also bin multiple C_ℓ's together which will have a symmetrizing effect. At $\ell > 30$ in temperature and polarization, we use the PLIK-LIKE bandpowers and covariance, rebinned as described above. We do not use polarization below $\ell < 30$ because the signal there is highly reionization-model dependent (e.g., Zaldarriaga, 1997). For the lensing reconstruction, we use the bandpowers and covariance from the "agressive" data cut. The window functions are shown in Fig. 7.2. One can see that the TE window functions are not strictly positive since they do not arise from an auto spectrum. For this reason, we cannot interpret them as a constraint on the amplitude of $P_{\rm m}(k,0)$, hence we show only TT and EE in Fig. 7.3. Although we do not do so here, one could interpret them as a constraint on a linear combination of the amplitude and derivative of $P_{\rm m}(k,0)$, however.

Cosmic shear

For cosmic shear, we use DES first-year constraints on the cosmic shear real-space two-point correlation functions $\xi_{\pm}^{ij}(\theta)$, where the i and j indices label different redshift bins (Troxel et al., 2017). These functions can be written in the form of equation (7.7),

$$\xi_{\pm}^{ij}(\theta) = \int d\ln k \, W_{\pm}^{ij}(\theta, k) P(k, 0), \tag{7.8}$$

where

$$W_{\pm}^{ij}(\theta, k) = \frac{1}{2\pi} \int_0^{\chi^H} d\chi \, \ell(\ell + 1/2) J_{0/4}(\theta\ell) \frac{q^i(\chi) q^j(\chi)}{\chi^2} \frac{P(k, \chi)}{P(k, 0)}, \tag{7.9}$$

the $q^i(\chi)$ are the lensing efficiency functions defined as usual (e.g. as in Troxel et al., 2017), and

$$k = \frac{\ell + 1/2}{\chi}. \tag{7.10}$$

We choose to bin together all of the redshift bins, producing a set of 5 fairly localized window functions for each θ bin, plotted in Fig. 7.2. Interestingly, one can see that ξ_+ produces window functions which are not strictly positive. This arises due to the weighting of the Bessel function inside of the integrand. Thus, similarly as for the CMB TE power spectrum, we do not plot these constraints on Fig. 7.3, although they could in theory also be interpreted as a joint constraint on the amplitude and derivative.

Galaxy clustering

For galaxy clustering, we use measurements of the halo power spectrum from a sample of luminous red galaxies from the SDSS DR7 (Reid et al., 2010). Using a model for the halo bias, we can relate these measurements to the underlying linear matter power spectrum in which we are interested. We use the model given in Reid et al. (2010) with free parameters b_0, a_1, and a_2. Fitting to our fiducial cosmological model, we find best-fit values of $1.24, 0.54$, and -0.33, respectively, at a pivot scale of $k_\star = 0.2\,\mathrm{Mpc} \cdot \mathrm{h}$.

7.1.3 Physical interpretation

Using the data and methods described above, we obtain a measurement of the 3D matter power spectrum at redshift $z = 0$ by combining different cosmological probes spanning four orders of magnitude in scales, from $k = 2 \times 10^{-4}$ to $k = 2\,\mathrm{h} \cdot \mathrm{Mpc}^{-1}$, and a wide range of cosmic history, from $z \sim 0$ to 1000, shown in Fig. 7.3. On scales of a few Mpc we use the latest Lyα forest data (SDSS-IV DR14) from which I computed the 1D power spectrum as explained in Chapter 4. I inferred the 3D matter power spectrum at $z = 0$ as explained in Sec. 7.1.1. On small scales, we also use cosmic shear real-space two-point correlation functions from the DES YR1 data release (Troxel et al., 2017). On scales of tens of Mpc, we use measurements of the halo power spectrum from a sample of LRGs from the SDSS seventh data release. For scales of hundreds of Mpc we use CMB data from Planck 2018 with temperature, polarization and lensing reconstruction power spectra measurements (Collaboration et al., 2018,?).

This provides a qualitative consistency test of the ΛCDM model. Although we do not perform any thorough quantitative tests, we have computed the χ^2 of the the data points shown in Fig. 7.3 against our fiducial model, ignoring any covariance between the data points, and using only the

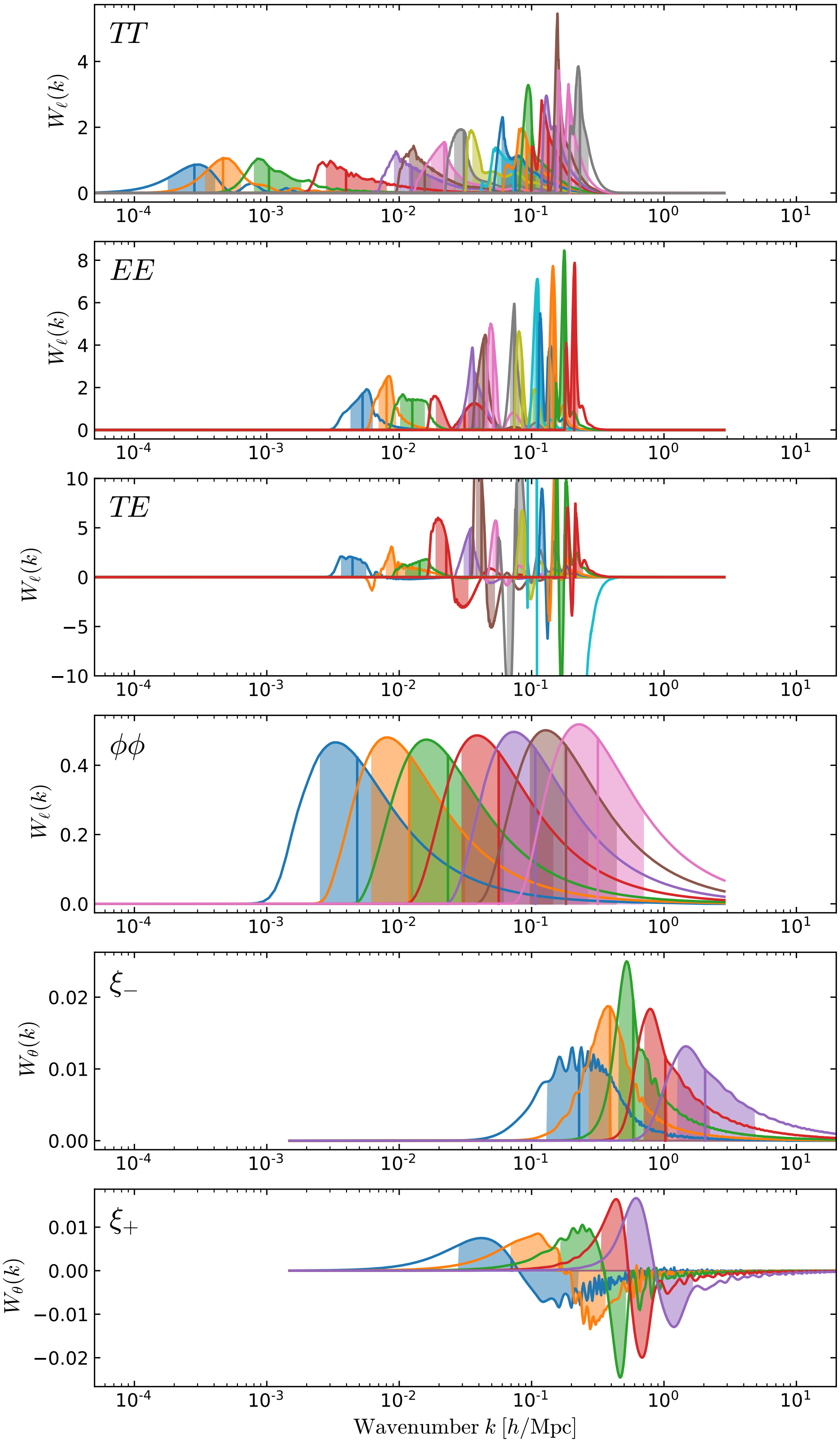

FIGURE 7.2 – *The window functions, $W_i(k)$, for several of the datasets considered here. The shaded region represents the middle 80% quantile of the absolute value of the function, which is the region denoted by the k-direction error bars in Fig. 7.3, and represents roughly to which k-scales a given data observation is sensitive to. Note that some observations have non-strictly positive windows, meaning we cannot interpret them as simply a measure of the overall amplitude of the matter power spectrum at a given scale, but rather some combination of this and its derivative.*

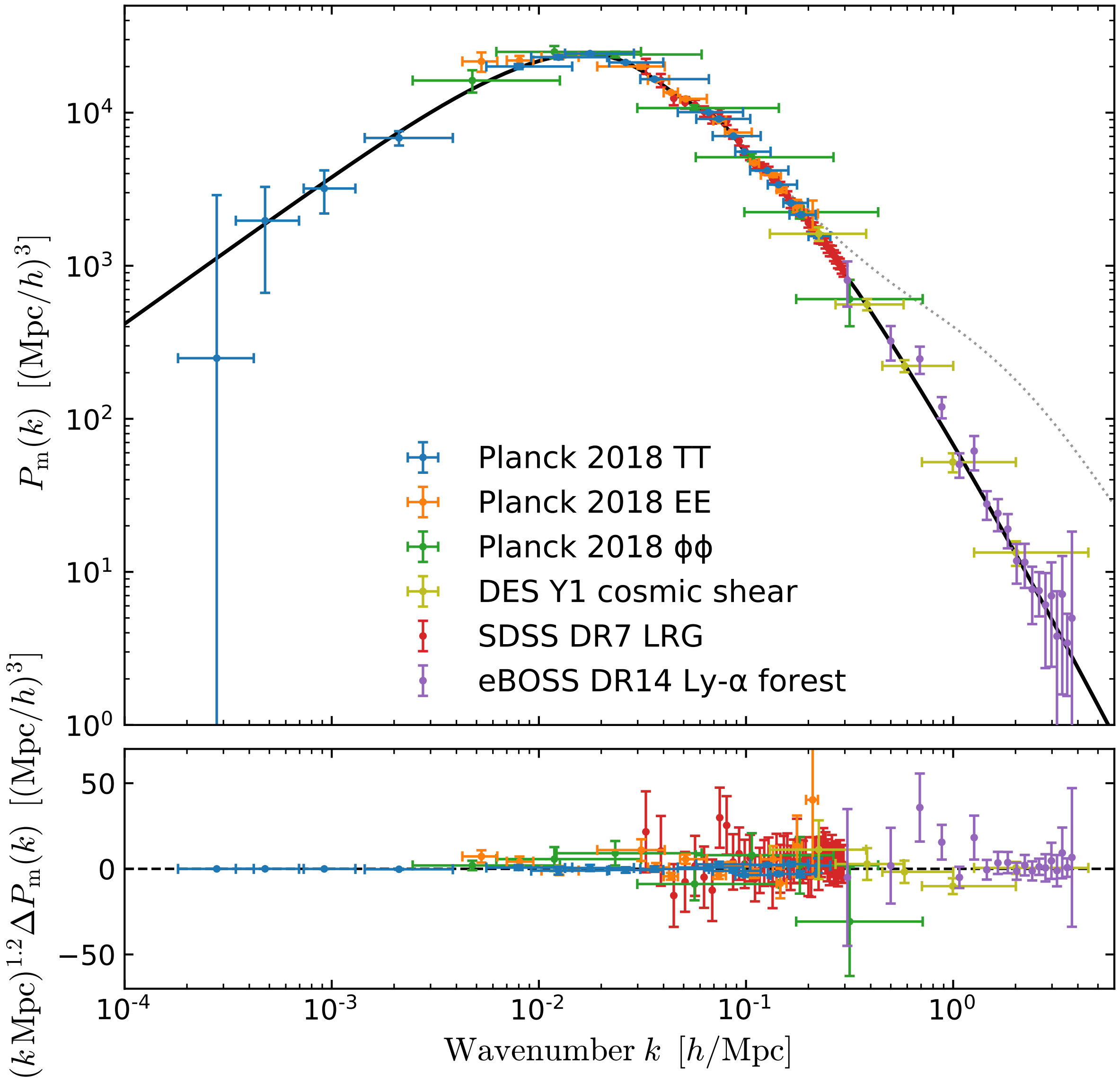

Figure 7.3 – **Top** : *Data points show inferences of the 3D linear matter power spectrum at* $z = 0$ *from Planck CMB data on the largest scales, SDSS galaxy clustering on intermediate scales, SDSS Lyα clustering and DES cosmic shear data on the smallest scales. In cases where error bars in the* k*-direction are present, we have used the method of Tegmark and Zaldarriaga (2002) to calculate a central 60% quantile of the region to which each data point is sensitive. In other cases, data points represent the median value of the measurement. The solid black line is the theoretical expectation given the best-fit Planck 2018 ΛCDM model (this model also enters the computation of the data points themselves). The dotted line for reference shows the theoretical spectrum including non-linear effects.* **Bottom** : *deviation of the data from the Planck best fit ΛCDM 3D matter power spectrum.*

error bars in the y-direction. We find $\chi^2 = 117.3$ for 108 degrees of freedom, which is consistent with an expected χ^2 fluctuation to within $1\,\sigma$. We stress that this number is only a very rough quantitative estimate of the consistency, but does at least highlight that no discrepancy is hiding in the residuals of Fig. 7.3. Our results thus highlight the good agreement of the ΛCDM model with observational data issued by independent experiments, covering a large range of cosmic times and cosmic scales.

7.2 Constraints on cosmological parameters

The previous section provides a qualitative study proving the consistency of the modern cosmological ΛCDM model across a a large range of cosmic times and cosmic scales, using different cosmological probes. In this part, I will focus on quantitative cosmological constraints using directly the Lyα signal alone or in combination with other probes. Chapter 2 shows that valuable information is brought by cosmological analysis of the 1D power spectrum of the Lyα forest since it probes small scales, unreachable by other probes. When cosmological signal is efficiently disantengled of that of astrophysical parameters, such as baryonic feedback or pressure smoothing, the Lyα forest is an invaluable tool at constraining the sum of the mass of the active neutrinos and at measuring the impact of the free streaming of keV-scale particles on the formation of structures. This section is then dedicated to the cosmological analysis of the sum of the mass of active neutrinos and the mass of thermal warm darm matter using the 1D power spectrum of the Lyα forest measurements derived in Chapter 4. This analysis led to an article published in Palanque-Delabrouille et al. (2020) to which I contributed in collaboration with my PhD advisor Nathalie Palanque-Delabrouille, the cosmology group at DphP and collaborators at RWTH Aachen University. I will mainly go through the steps I contributed to.

The outline of this section is as follow. The different data sets used to perform these cosmological analyses are presented in Sec. 7.2.1.1. Sec. 7.2.1.2 describes the methodology to construct the 1D power spectrum from hydrodynamical simulations and presents the additional parameters to account for imperfections in the modeling and measurements. Sec. 7.2.1.3 explains the frequentist methodology uded to statistically fit the data to theoretical predictions. Sec. 7.2.2 outlines cosmological constraints derived with Lyα data alone, and in particular the impact of the new AGN feedback parametrization. In Sec. 7.2.3 we discuss a mild-tension between Lyα and CMB data. Sec. 7.2.4 we combine Lyα and CMB data to constrain the mass of neutrinos under different assumptions for the underlying cosmological model to test the robustness of our results. Finally, Sec. 7.2.5 presents the thermal warm dark matter mass bounds using Lyα data alone.

7.2.1 Methodology and data sets

7.2.1.1 Data sets

Lyα

We use the 1D Lyα flux power spectrum measurement based on the DR14 BOSS and eBOSS data derived in Chapter 4. We perform the analysis on 435 data points, spread evenly over 35 bins in k space (from $k = 0.0011$ (km/s)$^{-1}$ to 0.0195 (km/s)$^{-1}$) and 13 bins in redshift space. We also explore an extended Lyα data set by considering the 1D Lyα power spectrum measurement from the one hundred XQ-100 quasars Yèche et al. (2017). These data cover three redshift bins at $z = 3.2$, 3.6, and 3.9. They exhibit a better resolution than the SDSS data, and thus allow an extension of the analysis to slightly higher k modes, namely to $k = 0.05$ (km/s)$^{-1}$ for $z = 3.2$, $k = 0.06$ (km/s)$^{-1}$ for $z = 3.6$, and $k = 0.07$ (km/s)$^{-1}$ for $z = 3.9$.

Cosmic Microwave Background

For the CMB data, we use the Planck 2018 data sets. The likelihoods are described in Aghanim et al. (2019), and the corresponding publicly available chains are used. We always use the full

Parameter	Definition
Ω_m	Matter fraction today (compared to critical density)
H_0	Expansion rate today in km $\mathrm{s}^{-1}\cdot\mathrm{Mpc}^{-1}$
σ_8	RMS matter fluctuation amplitude today in linear theory
n_s	Scalar spectral index
$\sum m_\nu$	Sum of neutrino masses
m_x	Mass of thermal relic WDM particle
α_s	Running of the power spectrum scalar index
$T_0(z=3)$	Normalization temperature of IGM at $z=3$
$\gamma(z=3)$	Logarithmic slope of δ-dependence of IGM temperature at $z=3$
$\eta^T(z<3)$	Logarithmic slope of z-dependence of T_0 for $z<3$
$\eta^T(z>3)$	Logarithmic slope of z-dependence of T_0 for $z>3$
η^γ	Logarithmic slope of z-dependence of γ
A^τ	Amplitude of the effective optical depth of Lyα absorption τ_{eff}
η^τ	Logarithmic slope of redshift dependence of τ_{eff}
$f_{\mathrm{Si\,III}}$	Fraction of Si III absorption relative to Lyα absorption
$f_{\mathrm{Si\,II}}$	Fraction of Si II absorption relative to Lyα absorption
z_{reio}	Redshift of reionization
A^{splice}	Amplitude of splicing correction
η^{splice}	Small-scale slope of splicing correction
A^{SN}	Amplitude of supernova feedback correction
A^{AGN}	Amplitude of AGN feedback correction
A^{UVfluct}	Amplitude of UV fluctuation correction
$A^{n,i}$	Amplitude of noise power correction for redshift bin i

TABLE 7.2 – *Definition of the parameters used sorted in three categories : cosmological, astrophysical, and nuisance.*

low-ℓ and high-ℓ temperature and polarization data, denoted as “P18”. When specified, we also include the Planck 2018 CMB lensing data, denoted as “lens.”.

Baryon acoustic oscillations

We also combine CMB data with measurements of the BAO scale by 6dFGS Beutler et al. (2011), SDSS main galaxy sample Ross et al. (2014), BOSS-LOWZ, and CMASS from DR12 Alam et al. (2017). Theses measurements are henceforth globally denoted “BAO”. The additional constraints that these measurement provide on cosmological parameters are included in the present work with their full correlation with CMB data. Both CMB and BAO constraints are taken from the Markov chains publicly available through the official Planck legacy archive.

7.2.1.2 Constructing theoretical predictions from simulations

To predict the 1D Lyα power spectrum for a given set of cosmological parameters, we use the three grids of hydrodynamical simulations covering the cosmological parameter space presented in Chapter 6 : the *G_BASE* grid corresponds the basic flat ΛCDM model, the *G_NU* grid explores the neutrino sector and the *G_WDM* grid explores the dark sector composed by hypothetical thermal relics. We also use an additional set of simulations to allow for the running of the scalar spectral index.

The likelihood is constructed upon three categories of parameters which are floated in the minimization procedure ; cosmological parameters, astrophysical parameters to model the IGM thermal model and nuisance parameters to describe imperfections in measurements and simulations. All parameter definitions are listed in Tab. 7.2, the next paragraph is dedicated to their full description.

The dependence of the power spectrum with cosmological and astrophysical parameters is modeled by a Taylor expansion of hydrodynamical simulation power spectra. The grid of simulations required for this interpolation consists of a reference simulation, called the best-guess simulation, centered on the Planck 2013 best-fit cosmology Ade et al. (2014) and an IGM thermal history in agreement with Meiksin (2009); Becker et al. (2011). The ensemble of parameters $\vec{\Theta}$ are in the central configuration $\vec{\Theta}_0$ for this best-guess simulation. So we can evaluate the dependence of the Lyα power spectrum around the best-guess configuration using a second-order Taylor expansion :

$$\begin{aligned} f\left(\vec{\Theta}_0+\Delta\vec{\Theta}\right) \simeq{}& f\left(\vec{\Theta}_0\right) \\ &+\sum_i \left.\frac{\partial f}{\partial\Theta_i}\right|_{\vec{\Theta}=\vec{\Theta}_0} \Delta\Theta_i \\ &+\frac{1}{2}\sum_i\sum_j \left.\frac{\partial^2 f}{\partial\Theta_i\partial\Theta_j}\right|_{\vec{\Theta}=\vec{\Theta}_0} \Delta\Theta_i\Delta\Theta_j \\ &+\mathcal{O}\left(|\Delta\vec{\Theta}|^3\right) \end{aligned} \tag{7.11}$$

Each parameter of the simulation grid require running $2n$ additional simulations for the $\pm\Delta\Theta_i$ first-order terms and $n(n-1)/2$ simulations for the second-order cross-terms where n is the number of parameters in the model. The dependence of the power spectrum with nuisance parameters is described through analytical corrections.

Cosmological parameters

The sets of simulations explore the $\{H_0,\Omega_m,n_s,\sigma_8,\sum m_\nu,1/m_x,\ \alpha_s\}$ cosmological parameter space. They cover the range $H_0 = 67.5\pm5$ km $\cdot$ s^{-1} $\cdot$ Mpc^{-1}, $\Omega_M = 0.31\pm0.05$, $n_s = 0.96\pm0.05$, $\sigma_8 = 0.83\pm0.05$. In all the runs we have $\Omega_b = 0.0221$.

Where the ΛCDMν model is assumed *G_NU* provides runs with $\sum m_\nu = 0.4$ or 0.8 eV and simulations for the cross-terms.

Where ΛWDM is assumed, we use additional runs from *G_ WDM* where the dark matter particles are thermal relics with masses $m_x = 2.5$ or 5.0 keV. I recall that the ΛWDM model is an exotic cosmological model where *all* the dark matter is made of WDM. We do not here consider models with mix of CDM and WDM. Since ΛCDM is reproduced for $m_x\to\infty$, we use $1/m_x$ instead of m_x in the Taylor expansion, and the simulations therefore probe $1\,\mathrm{keV}/m_x = 0$, 0.2 and 0.4.

When running of the scalar index is assumed, we use a set of simulations where the running parameter $\alpha_s \equiv dn_s/d\ln k$ is fixed to ±0.04. This allows us to introduce running directly in the Taylor expansion and thus fully account for its impact on the 1D Lyα flux power spectrum, whether direct or through cross-correlations with the other parameters.

Astrophysical parameters

The astrophysical parameters are chosen to follow the evolution of the IGM thermal state.

The temperature-density relation is modeled by a power law with

$$T(\delta, z) = T_0(z)\delta^{\gamma(z)-1}. \tag{7.12}$$

The evolution with redshift is modeled by a broken power-law for the amplitude T_0 at $z = 3$, defined with $T_0(z = 3) = 14000 \pm 7000\,\mathrm{K}$, the amplitude at $z = 3$, and the two logarithmic slopes $\eta^{T_0}(z < 3)$ and $\eta^{T_0}(z > 3)$. The evolution with redshift of the temperature density index γ is modeled by a simple power law with $\gamma(z = 3) = 1.3 \pm 0.3$, the amplitude at $z = 3$ and the logarithmic slope η^{γ}. Therefore, the set of astrophysical parameters to describe the IGM thermal evolution is composed by a total of 5 parameters $\{T_0$ (z = 3), $\eta^{T_0}(z < 3)$, $\eta^{T_0}(z > 3)$, γ, $\eta^{\gamma}\}$.

The photo-ionization rate of each simulation was fixed at each redshift to follow the empirical law $\tau_{\mathrm{eff}}(z) = A^{\tau}(1+z)^{\eta^{\tau}}$ with $A^{\tau} = 0.0025 \pm 0.0020$ and $\eta^{\tau} = 3.7 \pm 0.4$ in agreement with (Meiksin, 2009).

We include two amplitudes to model correlated absorptions by Lyα with Si II and Si III, as described in Sec. 7.1.1 (see equation (7.2)). The two amplitudes are used as multiplicative corrections to the flux power spectrum.

When ΛWDM is assumed, it requires to be able to lift the degeneracy between Jeans smoothing and WDM free streaming. Therefore, for WDM studies, we account for changes in the redshift of reionization by adding an astrophysical parameter that reproduces the impact of z_{reio} on the redshift and mode dependence of the flux power spectrum, as was done in Baur et al. (2016).

Nuisance parameters

Nuisance parameters allow us to account for uncertainties or corrections related to the measurements and imperfect modeling in the simulations.

For uncertainties in the measurements, we add the $A^{n,i}$ parameter for imperfect modeling of the noise (see Sec. 4.2.3 of Chapter 4).

On the theoretical side, simulations cannot include all the necessary physics to model the Lyα forest with the sufficient precision because of the limited computational power. First, all our grids of simulations are run without AGN and stellar feedbacks, which have large impact on the power spectrum, as seen in Chapter 6. The suppression of the power spectrum induced by AGN feedback is modeled using the analytical corrections derived in Sec. 6.2. The SN feedback acts on similar scales, and partially compensates the effect of AGNs. We adopt the study from Viel et al. (2013) to correct for the SN feedback. For both feedbacks, we apply a Gaussian prior around the central value of the correction.

The grid simulations contain the equivalent of 3072^3 particles of each type. Following a method originally suggested in McDonald (2003), this resolution is obtained by splicing together large-volume and high-resolution simulations, using a transition simulation that corrects the large-box simulation for its lack of coupling between small and large modes, and the high-resolution simulation for its small volume. The accuracy of the splicing technique is studied Palanque-Delabrouille et al. (2015) and Palanque-Delabrouille et al. (2015), and we correct for residual biases by the nuisance parameters A^{splice} and η^{splice}.

Finally, fluctuations in the intensity of the ionizing background (also referred to as UV fluctuations) are accounted for by an additive correction proportional to the measured power spectrum at the pivot wavenumber $k = 0.009\ (\mathrm{km/s})^{-1}$, as motivated by the study of Gontcho et al. (2014).

7.2.1.3 Frequentist methodology

In Palanque-Delabrouille et al. (2020) two different interpretation methodologies are used to assess the robustness of the results ; frequentist and Bayesian. I will only present results derived using the frequentist methodology since the two interpretations yield fully compatible results. Thus, this section is only dedicated to the description of the frequentist approach.

The determination of the coverage intervals of unknown cosmological parameters is based on the "classical" confidence level method originally defined by Neyman (1937). We start with the likelihood $\mathcal{L}(x, \sigma_x; \Theta)$, for a given cosmological model defined by the n cosmological, astrophysical and nuisance parameters $\Theta = (\theta_1, \ldots, \theta_n)$, and for data measurements x with Gaussian experimental errors σ_x. In the following, we adopt a χ^2 notation, which means that the following quantity is minimized :

$$\chi^2(x, \sigma_x; \Theta) = -2\ln(\mathcal{L}(x, \sigma_x; \Theta)) \ . \tag{7.13}$$

We first determine the minimum $\chi^2_{\min}$ of $\chi^2(x, \sigma_x; \Theta)$ leaving all the cosmological parameters free. Then, to set a confidence level (CL) on any individual cosmological parameter θ_i, we scan the variable θ_i : for each fixed value of θ_i, we minimize again $\chi^2(x, \sigma_x; \Theta)$ but with $n-1$ free parameters. The χ^2 difference, $\Delta\chi^2(\theta_i)$, between the new minimum and $\chi^2_{\min}$, allows us to compute the CL on the variable, assuming that the experimental errors are Gaussian,

$$\mathrm{CL}(\theta_i) = 1 - \int_{\Delta\chi^2(\theta_i)}^{\infty} f_{\chi^2}(t; N_{dof})dt \ , \tag{7.14}$$

with the χ^2 distribution

$$f_{\chi^2}(t; N_{dof}) = \frac{e^{-t/2} t^{N_{dof}/2-1}}{\sqrt{2^{N_{dof}}}\Gamma(N_{dof}/2)} \ , \tag{7.15}$$

where Γ is the Gamma function and the number of degrees of freedom N_{dof} is equal to 1. This profiling method can be easily extended to two variables. In this case, the minimizations are performed for $n-2$ free parameters and the confidence level $\mathrm{CL}(\theta_i, \theta_j)$ is derived from equation (7.14) with $N_{dof} = 2$.

In the analysis we also combine the χ^2 derived from the Lyα likelihood with that of Planck. In the frequentist analysis, we do not use the Planck likelihoods directly, but we use the Markov chains available in the official Planck[5] repositories instead. For instance, for the Planck 2018 TT+TE+EE configuration with massive neutrino, we take the chains from the directory `base_mnu/plikHM_TTTEEE_lowl_lowE`, which we reduce to the cosmological parameters $\{\sigma_8, n_s, \Omega_m, H_0, \sum m_\nu\}$ that are relevant for our analysis. The distribution of the chain elements allow us to estimate the posterior probability distributions for each parameter and the correlations between parameters. The flat positive prior applied to $\sum m_\nu$ causes a distortion of all posterior probability distributions, in particular for $\sum m_\nu$, and thus also for $\{\sigma_8, \Omega_m\}$, which are strongly correlated with $\sum m_\nu$. The posterior probability distribution becomes asymmetric and cannot be modeled by a simple Gaussian distribution. To account for such effects, we first apply a Principal Component Analysis on the reduced chain that allows us to determine the linearly uncorrelated variables, called the principal components. We then model the distribution of each principal component by several asymmetric Gaussians. This strategy is validated

---

[5] `https://wiki.cosmos.esa.int/planck-legacy-archive/index.php/Cosmological_Parameters`

	DR14		DR9	
	$\sum m_\nu = 0$	Varying $\sum m_\nu$	$\sum m_\nu = 0$	Varying $\sum m_\nu$
T_0 (z=3) (10^3K) .	8.5 ± 1.9	8.5 ± 2.0	9.5 ± 3.5	9.5 ± 3.5
γ	0.93 ± 0.14	0.93 ± 0.14	1.0 ± 0.2	1.0 ± 0.2
σ_8	0.826 ± 0.020	0.826 ± 0.021	0.830 ± 0.00322	0.830 ± 0.0032
n_s	0.954 ± 0.006	0.954 ± 0.006	0.939 ± 0.010	0.939 ± 0.010
Ω_m	0.269 ± 0.009	0.269 ± 0.009	0.293 ± 0.014	0.293 ± 0.014
$\sum m_\nu$ (eV , 95% CL)	-	< 0.58	-	< 1.1

Table 7.3 – *Preferred astrophysical and cosmological parameter values (68.3% confidence level) for the* ΛCDM *and* ΛCDMν *models, for DR14 and DR9 Lyα data combined with a Gaussian prior* $H_0 = 67.3 \pm 1.0\,\mathrm{km \cdot s^{-1} \cdot Mpc^{-1}}$.

in Palanque-Delabrouille et al. (2020) for a few configurations by comparing the limits obtained on $\sum m_\nu$ with this modeling with the limits derived directly from the bayesian approach using the full likelihood. The agreement between the two approaches was typically at the level of a few percent.

7.2.2 Constraints with Lyα data alone

In this section, I first focus on the constraints from Lyα data alone and I estimate the impact of the new AGN feedback parametrization derived in Chapter 6 proving its necessity to get robust scientific interpretation from Lyα analysis.

7.2.2.1 Comparison between DR9 and DR14

As was noted in Palanque-Delabrouille et al. (2015), the Lyα forest power spectrum only weakly depends on the Hubble parameter, and is unable to constrain H_0 by itself. When using Lyα data alone, we thus adopt the same Gaussian prior constraint as in Palanque-Delabrouille et al. (2015); Baur et al. (2016, 2017); Chabanier et al. (2019), which is taken from the Planck 2015 TT+lowP results Planck Collaboration et al. (2015), namely $H_0 = 67.3 \pm 1.0\,\mathrm{km \cdot s^{-1} \cdot Mpc^{-1}}$. Note that most recent result from Planck Collaboration et al. (2018b) using TT, TE, EE, low E + lensing gives $H_0 = 67.36 \pm 0.54$, in perfect agreement with the prior mentioned above. In Palanque-Delabrouille et al. (2015), we explicitly checked that bounds on other parameters depend very weakly on the choice of H_0 prior. In particular, the Lyα posteriors are *not* significantly different when we combined the Lyα data with an H_0 prior taken from the SH$_0$ES results Riess et al. (2019).

The IGM thermal history is one of the main sources of nuisance in this study. To encompass a large range of possible histories, we marginalize over the thermal parameters of Tab. 7.2 to derive constrains either on $\sum m_\nu$ (Sec. 7.2.4) or on the mass of a WDM thermal relic (Sec. 7.2.5). Because of the freedom allowed in the modeling, the uncertainties on the thermal parameters are large, and the 2σ range on T_0, γ and mean flux overlaps with observational measurements. To further test the robustness of our result, we checked that by imposing different thermal models (e.g. one in agreement with Becker et al. (2011)) has little impact on the measured bounds.

	noAGN	weakAGN	fidAGN	strongAGN
σ_8	0.82 ± 0.02	0.83 ± 0.02	0.83 ± 0.02	0.83 ± 0.02
n_s	0.958 ± 0.005	0.950 ± 0.005	0.949 ± 0.005	0.946 ± 0.005
Ω_m	0.268 ± 0.009	0.269 ± 0.009	0.270 ± 0.009	0.269 ± 0.009
$T_0(z=3)$ (10^3 K)	8.5 ± 2.0	8.6 ± 1.8	8.64 ± 1.9	8.7 ± 1.2
γ	0.92 ± 0.13	0.95 ± 0.12	0.93 ± 0.14	0.97 ± 0.15
A^τ (10^{-3})	2.33 ± 0.06	2.37 ± 0.06	2.38 ± 0.06	2.40 ± 0.06
η^τ	3.83 ± 0.03	3.83 ± 0.03	3.84 ± 0.03	3.84 ± 0.03

TABLE 7.4 – *Best-fit values and 68% confidence levels of the cosmological and astrophysical parameters for the* ΛCDM *model when imposing the AGN feedback correction and not letting it vary in the fit. In the first column no correction for the AGN feedback is applied, we apply the upper bound (weakAGN), the fiducial (fidAGN) and the lower bound (strongAGN) correction in the second, third and fourth columns respectively.*

Tab. 7.3 displays the best-fit values for the ΛCDMν model using Lyα data alone, either with DR9 data or with DR14 data (i.e. measurements derived in this thesis). The flux power spectrum of the best-fit model is shown in Fig. 7.4. Compared to the DR9 analysis, there is a noticeable improvement in the agreement between the data and the best-fit model, in particular for high k and high z. The best fit on Lyα data alone slightly differs from the one obtained on the DR9 data for two cosmological parameters : n_s, which increased from 0.938 ± 0.010 in Palanque-Delabrouille et al. (2015) to 0.954 ± 0.006, and Ω_m, which decreased from 0.293 ± 0.014 to 0.269 ± 0.009. We investigated the origin of this $\sim 1.5\,\sigma$ shift. Restricting the eBOSS data to the forests in common with those from DR9, we measure $n_s = 0.945 \pm 0.008$ and $\Omega_m = 0.278 \pm 0.015$. To further mimic the DR9 selection, we then replace the automated catalogs of BAL quasars and DLAs by the visual catalogs that were used in DR9. The fit on the resulting sample gives $n_s = 0.935 \pm 0.008$ and $\Omega_m = 0.282 \pm 0.015$, in good agreement with the cosmology obtained with the analysis of Palanque-Delabrouille et al. (2015) on the DR9 sample. The change in the data is therefore at the origin of the small shift in best-fit cosmological parameters.

7.2.2.2 Impact of the new AGN feedback parametrization

To highlight the impact of the AGN feedback parametrization derived in Chapter 6, we compute cosmological parameters in the four following situations : without applying AGN correction or applying the three corrections that span the AGN feedback parametrization uncertainty interval, i.e. the upper (weakAGN), fiducial (fidAGN) and lower (strongAGN) corrections

Thus, for this section only, we fix the nuisance parameter A^{AGN} for AGN feedback and do not let it free in the fit. In practice, we directly correct our theoretical predictions with the functions given in Tab. 6.4 of Chapter 6. Here, I aim at being illustrative to show the impact of the fiducial correction on the first hand, and, on the other hand, the variation of the cosmological prameters on the AGN feedback parametrization uncertainty range.

Tab. 7.4 presents the best-fit values in the ΛCDM model for the four cases : without AGN correction, with the weakAGN, fidAGN or strongAGN corrections. The most impacted cosmological parameter is the scalar spectral index n_s. I show in Fig. 7.5 the inferred values n_s for the four configurations. It is an expected result as AGN feedback tends to increase the slope of the flux power spectrum, it is therefore degenerate with n_s. Not taking into account AGN feedback yields a bias of about 1% which represents two standard deviations of n_s. However, it varies on

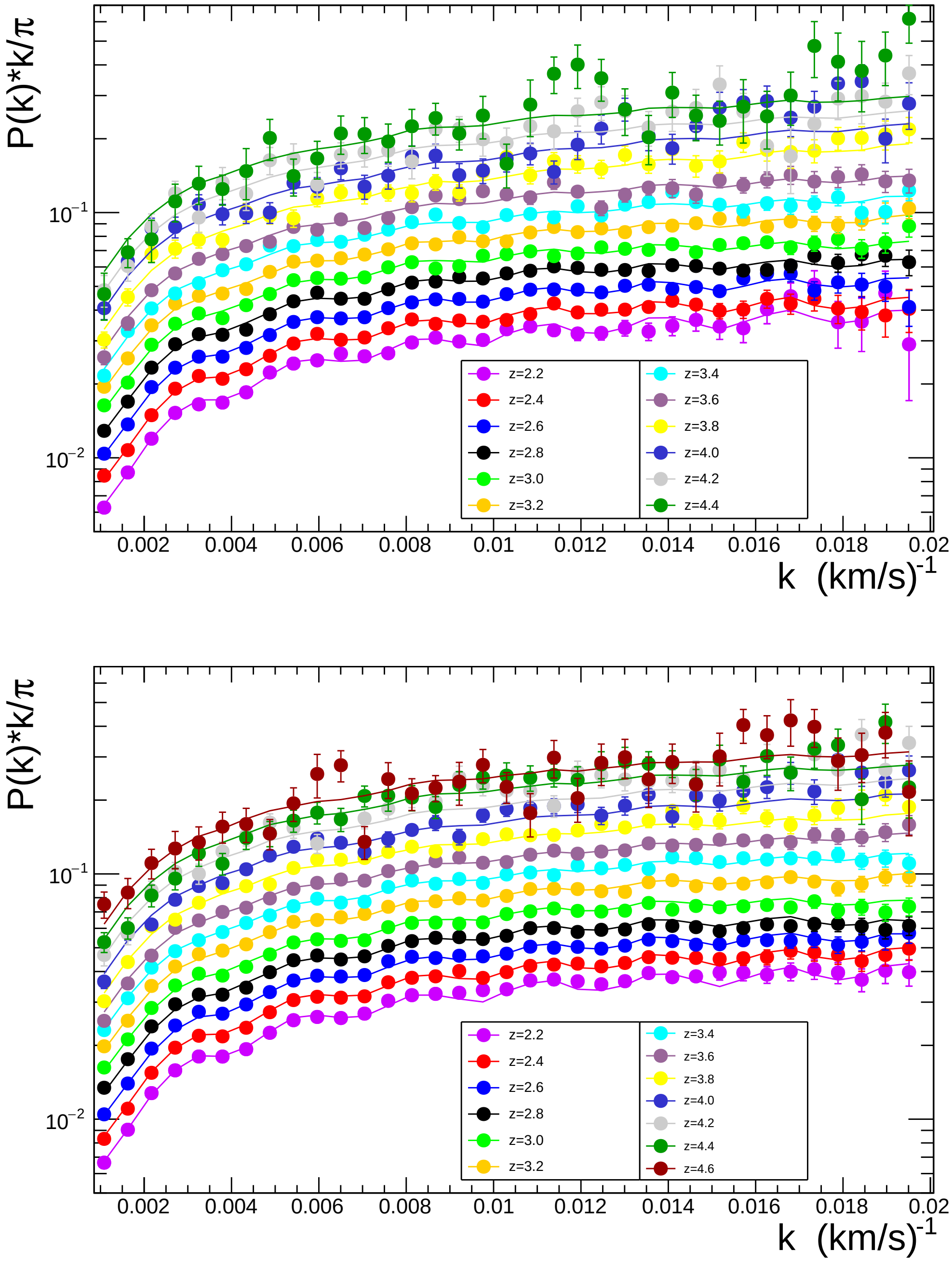

FIGURE 7.4 – ***Top :*** *BOSS DR9 Lyα data from Palanque-Delabrouille et al. (2015) and best-fit* ΛCDMν *model.* ***Bottom :*** *eBOSS DR14 Lyα data from this thesis and best-fit* ΛCDMν *model.*

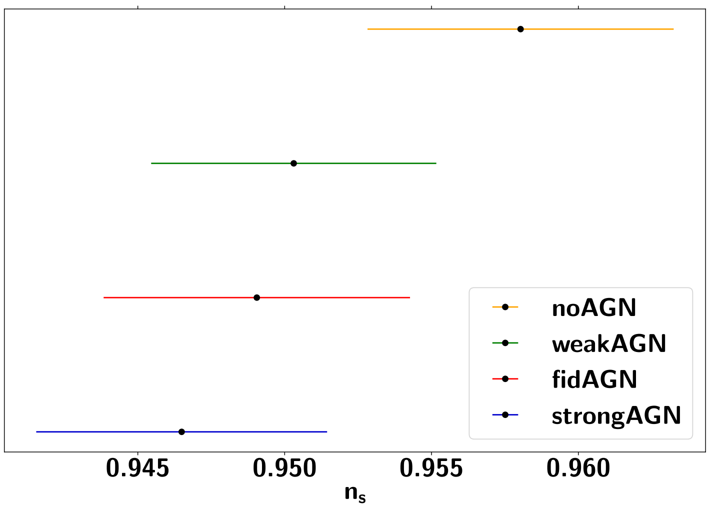

FIGURE 7.5 – *Scalar spectral index bias induced by AGN feedback on the 1D Lyα power spectrum in the ΛCDM model. We show the inferred values for the four following cases : no AGN correction, the weakAGN, fidAGN and strongAGN corrections.*

less than 0.5% on our uncertainty range in spite of the large suppression range allowed at low redshift.

The bias reaches 2% for σ_8, but because it is less constrained, the shift is contained within the statistical error. Finally Ω_m does not present any significant deviation.This is also expected because, as shown in Fig. 9 of Borde et al. (2014), varying Ω_m impacts the formation of small-scale structures, hence it has more significant impact at large k's.

The astrophysical parameters are more impacted with biases of 1% and from 3% to 6% for T_0 and γ respectively. However it stays contained in the uncertainty range as the IGM thermal state is not well constrained by medium-resolution SDSS data. Also, the AGN feedback increases the mean flux with a shift from 2% to 3% on the effective optical depth amplitude which represents one standard deviation.

Finally the uncertainties on cosmological and astrophysical parameters are hardly impacted by the AGN feedback. The uncertainty on the amplitude temperature at $z = 3$ decreases from 2.0 to 1.2 when applying the strongAGN correction. However, the thermal history is described by a total of seven, parameters showing significant correlations, e.g. 75% for $T_0(z = 3)$ and $\gamma(z = 3)$. While the uncertainty on $T_0(z = 3)$ is decreased, the uncertainties on $\gamma(z = 3)$ and on the redshift evolutions are all increased, which mitigates the conclusion on the precision of the estimated temperature.

7.2.3 Mild tension between Lyα and CMB data

Before combining the Lyα data with CMB data, we compare the results from the Lyα power spectrum (see Tab. 7.3) and from Planck 2018 (see Tab. 7.5) with minimal assumptions on the cosmological model, i.e. in the framework of the ΛCDMν model with a free value of $\sum m_\nu$. The common free parameters in the Lyα and Planck likelihoods are the primordial spectrum amplitude and spectral index $\{\sigma_8$ or $A_s, n_s\}$, the fractional density of matter Ω_m, and possibly the neutrino mass $\sum m_\nu$. There are two more common parameters Ω_b and H_0, but the Lyα

	P18	P18 + lens. +BAO
σ_8	0.804 ± 0.018	0.815 ± 0.009
n_s	0.9630 ± 0.0048	0.966 ± 0.004
Ω_m	0.321 ± 0.014	0.310 ± 0.007
$100\Omega_b$	2.232 ± 0.016	2.241 ± 0.014
H_0 $(\mathrm{km \cdot s^{-1} \cdot Mpc^{-1}})$	66.9 ± 1.1	67.81 ± 0.5
$\sum m_\nu$ (eV , 95% CL)	< 0.286	< 0.113

TABLE 7.5 – *Preferred cosmological parameter values (68.3% confidence level) fo the* ΛCDMν *model, for Planck data alone or combined with BAO data.*

data are so weakly sensitive to them that we fixed Ω_b and imposed an H_0 prior that guarantees agreement with Planck.

For all common parameters but one, we find excellent agreement between the confidence bounds derived from Lyα data and CMB data. This can be checked directly from Tab. 7.3 and Tab. 7.5, or visually by comparing two-dimensional contours in the $\{\sum m_\nu, \sigma_8, n_s\}$ plane in Fig. 7.6 showing the Lyα contours in red and the P18 contours in blue. For instance, for n_s, the Lyα and P18+lens+BAO bounds are compatible at the 1.4σ level.

We find a mild tension between the Ω_m values derived from CMB data ($\Omega_m \sim 0.31$) and Lyα data ($\Omega_m \sim 0.27$). The tension is present with or without massive neutrinos, and with respect to both the P18 and P18+lens+BAO datasets. It reaches a 3.6σ tension between the Ω_m values derived from Lyα data and the P18+Lens+BAO combination. This tension is displayed in the $\{\Omega_m, n_s\}$ plane in the left panel of Fig. 7.7.

Fig. 7.7 (left panel) shows that for a fixed value of Ω_m around 0.31, this mismatch could instead be interpreted as a mild tension on n_s. This is reminiscent of a similar tension on n_s found with previous Lyα data from BOSS DR9 discussed in Palanque-Delabrouille et al. (2015). When fitting the parameters of the ΛCDM model or its extensions to Lyα data, n_s and Ω_m are always anti-correlated, because they both affect the overall slope of the flux power spectrum in the same direction. Therefore the tension on Ω_m in the present version of the data set and likelihood is likely to have the same origin as the tension on n_s in the previous version : whether the tension is interpreted as one on n_s or as one on Ω_m strongly depends on the modeling of the data and its systematics. In any case, since the analysis presented in this work is based on the most up-to date data set and on the most advanced systematic modeling of the BOSS and eBOSS flux power spectrum, we will concentrate on the Ω_m tension and its possible origins.

We stress that the Lyα results are nicely consistent with those from weak lensing (WL) surveys. Over the past years, there has been a mild but persistent tension between likelihood contours in the (Ω_m, σ_8) plane inferred from Planck data and from WL surveys, when assuming either a ΛCDM or ΛCDMν cosmology. This is commonly referred to as the "σ_8 tension", although Ω_m is also involved. The tension is actually best seen when quoting results on the combination $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5}$ which is orthogonal to a direction of degeneracy in the WL posteriors.

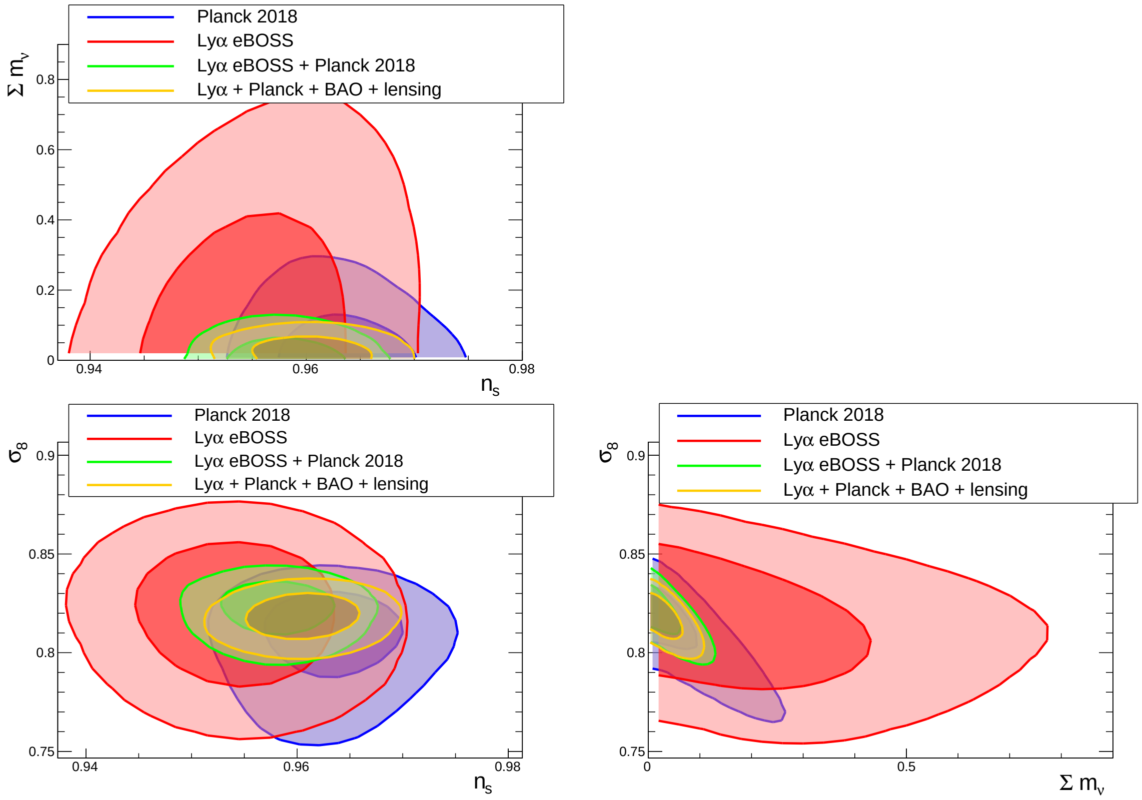

FIGURE 7.6 – *2D isocontours in the sub-space* $\{\sum m_\nu,\ \sigma_8,\ n_s\}$ *for the* ΛCDMν *model and various combinations of CMB, BAO and Ly*α *data. We show the* 68% *and* 95% *contours.*

In Fig. 7.8, we show the contours of the Lyα data in the (Ω_m, σ_8) and (Ω_m, S_8) planes, compared with those from one of the most recent joint analyses Asgari et al. (2019) of several WL data sets (DES-Y1 Drlica-Wagner et al. (2018); Zuntz et al. (2018), KV450 Wright et al. (2018)), and finally compared with Planck contours, assuming in each case a ΛCDM cosmology. The left panel of Fig. 7.8 shows that the σ_8 tension can be equally well interpreted as an Ω_m tension. While the CMB versus WL tension is strongest in the S_8 direction, the CMB versus Lyα tension is strongest in the Ω_m direction. It is striking to see that WL and Lyα data, which are two late time probes of a similar range of scales, agree with each other at the 1σ level, while they are both in tension with the Planck best-fit ΛCDM model at the 2.5σ to 3.6σ level.

In Palanque-Delabrouille et al. (2020) a full systematic search of the origin of this tension was performed, investigating possible sources of systematics both at the level of the modeling of the Lyα 1D flux power spectrum and at the level of the Lyα data analysis. Sub-samples of the data where considered with cuts on scales or redshifts. On the modeling side, the model used to fit the Lyα data was modified in several ways. For instance, a correction for the incompleteness of the masking of DLAs was used instead of the systematics derived in Chapter 4, or new correction models for the bias produces by the splicing technique were implemented. But none of these tests yielded any significant change in the Lyα best-fit cosmological values.

---

[6] Instead of using the DES-Y1 redshift distributions Troxel et al. (2017), these new analyses use photometric redshifts from COSMOS-2015 Joudaki et al. (2019); Asgari et al. (2019); Laigle et al. (2016). Note also that we are considering a flat ΛCDM model in agreement with the cited analyses.

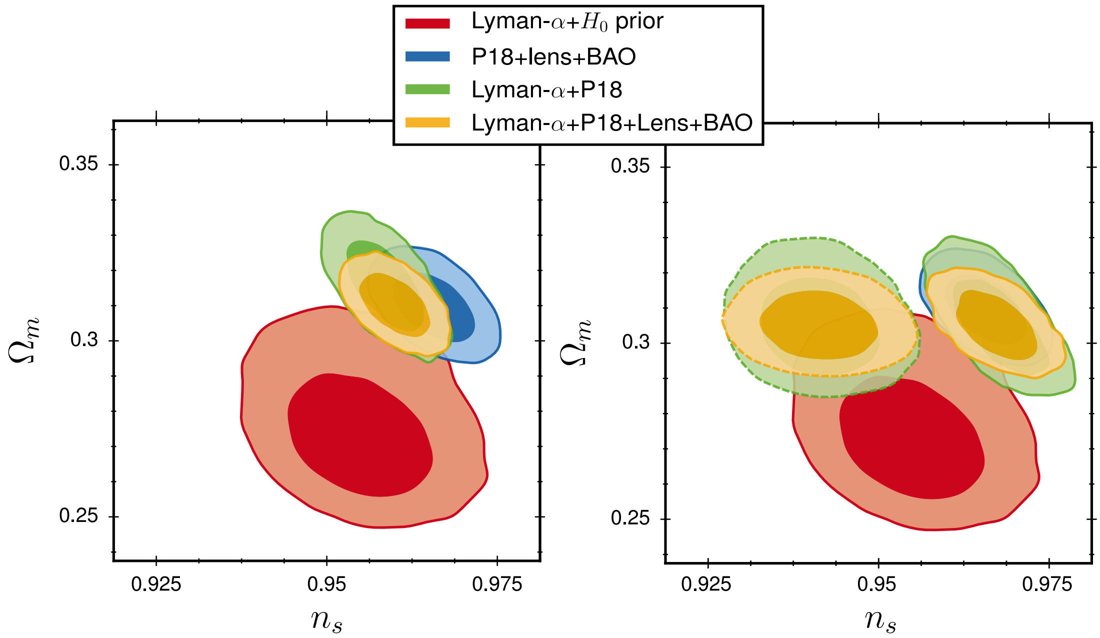

FIGURE 7.7 – *2D isocontours in the sub-space* $\{\Omega_m,\ n_s\}$ *assuming various cosmological models and combinations of CMB, BAO and Ly*α *data. We show the* 68.3%(1σ) *and* 95.4%(2σ) *limits.* **Left :** ΛCDMν *model, showing a mild tension.* **Right :** ΛCDMν *model with two independent tilts for the Planck and Ly*α *likelihoods.*

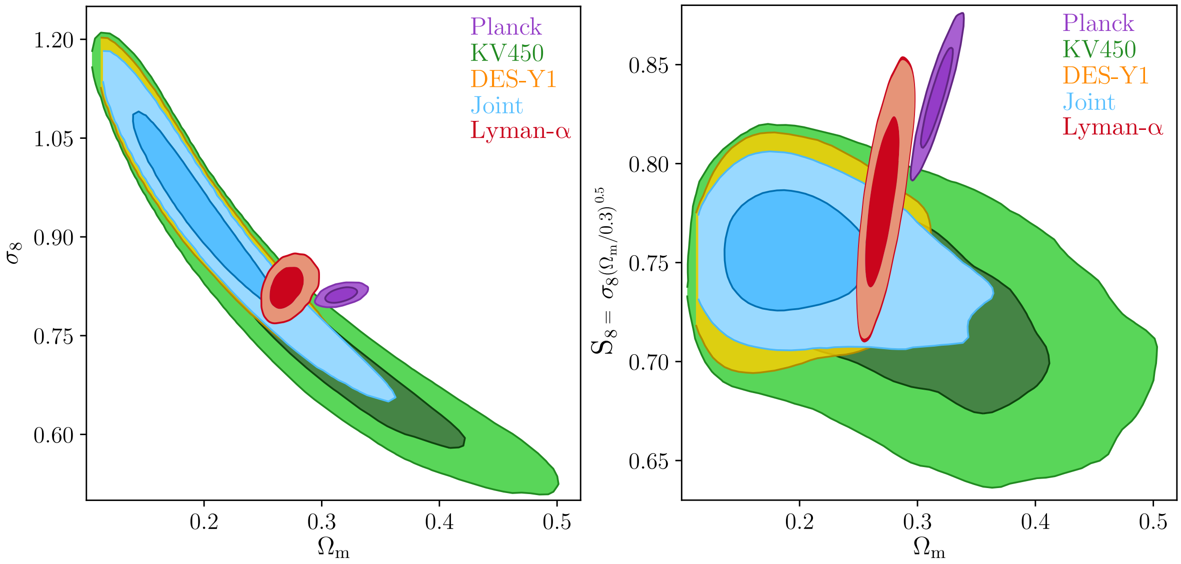

FIGURE 7.8 – *Comparison of the* $\{\sigma_8, \Omega_m\}$ *and* $\{S_8, \Omega_m\}$ *planes for Planck, Weak Lensing surveys (DES, KiDS+VIKING), and Ly*α *data, where* $S_8 \equiv \sigma_8\,(\Omega_m/0.3)^{0.5}$. *The tension between Ly*α *and CMB data is here best described as a tension in* Ω_m. *The COSEBI-based redshift-recalibrated analyses for DES-Y1*[6], *KV450, and their joint constraint are taken from Asgari et al. Asgari et al. (2019).*

We cannot exclude the possibility that the tension between the value of Ω_m preferred by CMB and Lyα data originates from yet another unidentified systematics (or imperfection in the computation of the Lyα likelihood). Indeed, with statistical uncertainties on the data points down at the percent level, results on cosmological parameters are now hitting the systematics floor. Despite the great care that went into their modeling, instrumental features – such as correction of spectrograph resolution and subtraction of noise power – affect the lowest redshift bins and the smallest scales at a level comparable to the statistical uncertainties. Uncertainties also arise on the simulation side, mostly in relation to the use of the splicing technique mentioned above. Although the bias induced by this approach was measured to be small, and although we mitigate the risk of an imperfect modeling by marginalizing over the parameters that correct for the impact of splicing, a residual bias on the large-scale correction would affect the slope of the 1D flux power spectrum and could be responsible for the observed tension. Finally, in spite of major efforts that were conducted to model with precision the AGN feedback effects and to take into account large variations allowed by different sub-grid modeling, we can neither reject the possibility of its imperfect modeling. As we saw in Sec. 7.2.2.2, AGN feedback impacts the inferred n_s value at 2%. Another road of improvement is the modeling of SN feedback which also affects the slope of the power spectrum and is modeled using the work from Viel et al. (2013). The latter uses SPH simulations with mass outflow rates at least one order of magnitude larger than observations, and could exaggerate the thermal effects of SN feedback in the low-density IGM.

7.2.4 Results on $\sum \mathbf{m}_\nu$

Lyα data alone have a sensitivity to $\sum m_\nu$ at the level of about 1 eV du to the fact that the scales probed by the Lyα forests are in the region where the ratio of the power spectra for massive to massless neutrinos is quite flat (see Fig. 1.9 of Chapter 2). Therefore, $\sum m_\nu$ presents a large degeneracy with σ_8 since both measure the amplitude of the Lyα power spectrum. This degeneracy can be broken and we can obtain a tight constrain by combining Lyα data with CMB data, which probe the initial power spectrum unaffected by the neutrino free streaming. However, to safely combine independent measurements in order to obtain joint constraints it is primordial that the constraints obtained taking each measurement separately are consistent. Therefore, to combine these two data sets we assume a basic extension of the cosmological model with a running of the primordial spectral index that may reduce this mild tension. We first present different methods to combine Lyα and CMB data and in particular the ΛCDMν+α_s cosmology, then we discuss active neutrino mass bounds and their robustness against different assumptions.

7.2.4.1 Combining Lyα and CMB data

CMB and Lyα data can be combined with different assumptions on the underlying cosmology.

First, one can adopt the point of view that the tension on Ω_m described in Sec. 7.2.3 is sufficiently small that it should not prevent us from combining the data sets while sticking to the ΛCDMν model. In this case, the combined limits on cosmological parameters are presented in Tab. 7.6. We will come back to the discussion of the neutrino mass bounds found in this analysis in a dedicated Sec. 7.2.4. Since the Planck data has more statistical weight, the Ω_m values in the combined fit are driven to Planck best-fit values, as shown in Fig. 7.7. The other cosmological parameters, already in good agreement between their best-fit values for the two data sets taken

	P18 + Lyα	P18 + Lyα +lens. +BAO
T_0 (z=3) (10^3K) .	9.7 ± 1.7	9.8 ± 2.0
γ	0.69 ± 0.10	0.68 ± 0.11
σ_8	0.825 ± 0.006	0.819 ± 0.008
n_s	0.958 ± 0.003	0.961 ± 0.003
Ω_m	0.311 ± 0.006	0.308 ± 0.006
$\sum m_\nu$ (eV , 95% CL)	< 0.099	< 0.089

Table 7.6 – *Preferred astrophysical and cosmological parameter values (68.3% confidence level) for the* ΛCDMν *model, for combined Ly*α*, CMB and BAO data.*

	P18 + Lyα	P18 + Lyα +lens. +BAO
T_0 (z=3) (10^3K)...	7.6 ± 1.9	7.6 ± 1.8
γ	0.88 ± 0.13	0.88 ± 0.08
σ_8	0.824 ± 0.008	0.820 ± 0.008
n_s(Planck).......	0.965 ± 0.004	0.968 ± 0.004
n_s(Lyα)	0.942 ± 0.006	0.942 ± 0.005
Ω_m	0.304 ± 0.010	0.304 ± 0.006
$\sum m_\nu$ (eV , 95% CL)	< 0.126	< 0.104

Table 7.7 – *Preferred astrophysical and cosmological parameter values (68.3% confidence level) for the* ΛCDMν *model, for combined Ly*α*, CMB and BAO data, when introducing artificially two distinct* n_s *value in the Ly*α*and CMB likelihood.*

	P18+Lyα	Lyα+P18 +lens. +BAO
T_0 (z=3) (10^3K)...	7.8 ± 1.8	8.0 ± 1.8
γ	0.80 ± 0.12	0.80 ± 0.12
σ_8	0.825 ± 0.007	0.821 ± 0.007
n_s	0.962 ± 0.003	0.962 ± 0.003
Ω_m	0.306 ± 0.007	0.307 ± 0.006
α_s	-0.010 ± 0.003	-0.010 ± 0.003
$\sum m_\nu$ (eV , 95% CL)	< 0.105	< 0.089

Table 7.8 – *Preferred astrophysical and cosmological parameter values (68.3% confidence level) for the* ΛCDMν+α_s *model, for combined Ly*α*, CMB and BAO data.*

individually, do not change significantly. While Planck data carry no information per se on the IGM thermal history, they can affect the best-fit values of the astrophysical parameters through their correlations with the cosmological terms. For instance, the anti-correlation between Ω_m and γ – and to a lesser extent between n_s and γ – is causing γ to decrease compared to its value for the fit with Lyα data alone, since both Ω_m and n_s are increased.

Since for a fixed value of Ω_m around 0.31, the tension could be interpreted as a tension in n_s, it is interesting to check whether the two data sets can be brought in better agreement by allowing the overall slope to be different for Lyα and CMB data. To check this, we first perform an academic study in which we assume different values of n_s in the CMB and in the Lyα likelihoods. Our results for that case are presented in Tab. 7.7.

In this configuration, the global χ^2 decreases by ~ 13.8 for P18+lens+BAO+Lyα compared to the ΛCDMν fit, which represents an improved compatibility between Lyα and P18+lens+BAO data. The corresponding two-dimensional $\{\Omega_m, n_s\}$ contours, displayed in the right panel of Fig. 7.7, now feature two distinct regions ; one (dashed lines) for n_s(Lyα) and another (solid lines) for n_s(Planck). The two values of n_s are centered around $n_s = 0.967$ for Planck and $n_s = 0.941$ for Lyα. In the right panel of Fig. 7.7, the improved compatibility shows up in the fact that the red contour (Lyα only) is now in better agreement with the green and yellow dashed contours (combined data set).

Of course, the improved agreement occurs at the price of rather unphysical assumptions. This exercise is however not completely artificial, because the CMB and Lyα data sets probe different fluctuations (photon perturbations or a tracer of baryons and CDM fluctuations) on different scales and at different times. Thus there could be many physical and sensible reasons for which the overall slope of the two observables are not correlated in the way predicted by the ΛCDM or ΛCDMν models.

First, the primordial power spectrum could have a different effective slope $\frac{d \ln \mathcal{P}_\mathcal{R}}{d \ln k}$ due to physical mechanisms taking place during inflation. For instance, a large curvature in the inflaton potential could produce a running of the spectral index, i.e. a continuous variation of $\frac{d \ln \mathcal{P}_\mathcal{R}}{d \ln k}$ with k Kosowsky and Turner (1995), while a kink in the potential could lead to a feature in the primordial spectrum with different spectral indices on large and small scale Joy et al. (2008).

Second, the growth rate of dark matter could be reduced on small scales during radiation and/or matter domination, for instance due to small interactions between dark matter and other species or self-interactions in the dark matter sector, or by a small departure from Einstein gravity. Since baryons fall in the gravitational potential wells of dark matter, this reduction would propagate to the baryons and to the flux power spectrum. An example of a mechanism leading to a small reduction of the effective slope of the matter power spectrum $\frac{d \ln \mathcal{P}_\mathrm{m}}{d \ln k}$ is provided by the scattering dark matter model of Buen-Abad et al. (2015). Another case is that of $f(R)$ gravity, that leads to a scale-dependent linear growth factor with less growth on small scales **?**.

In principle, a dedicated analysis would be needed in order to investigate up to which extent each of these models can reduce the mild tension between Lyα and CMB data. Here, we limit ourselves to the most studied among the previous models, featuring a running of the scalar index, $\alpha_s = \frac{d^2 \ln \mathcal{P}_\mathcal{R}}{d \ln k^2} = \frac{d n_s}{d \ln k}$, treated as constant over the range of scales probed by both CMB and Lyα data.

On the theoretical side, the running of the spectral index is usually connected to the physics of inflation, but we should keep in mind that it could be seen as an effective parametrization of some of the other models described previously – in particular, a negative α_s gives a reduction of the amplitude and of the effective slope of the small-scale matter power spectrum that could mimick a scale-dependent reduction of the linear growth factor. The simplest inflationary models predict that the running of the spectral index should be of second order in inflationary slow-roll parameters and therefore small, $|\alpha_s| \sim (n_s - 1)^2 \sim 10^{-3}$ Kosowsky and Turner (1995). Nevertheless, it is possible to accommodate a larger scale dependence of n_s by adjusting the third derivative in the inflaton potential (see for instance Kobayashi and Takahashi (2011); McAllister et al. (2010)).

On the experimental side, recent CMB experiments have a mixed history of null results and a-few-sigma detections of running of the scalar index. The final 9-year WMAP analysis found no evidence of running using WMAP alone, with $\alpha_s = -0.019 \pm 0.025$ at 68% CL, while the combination of WMAP data with the first data releases from ACT and SPT found a negative running at nearly the 2σ level with $\alpha_s = -0.022 \pm 0.012$ Hinshaw et al. (2013). The ACT 3-year release measured $\alpha_s = -0.003 \pm 0.013$ when combining with WMAP-7 Sievers et al. (2013). A negative running was detected at just over 2σ by SPT, $\alpha_s = -0.024 \pm 0.011$ Hou et al. (2014). The Planck 2018 results, while roughly consistent with zero running of the scalar spectral index, indicate a $\sim 1\sigma$ preference for negative running, $\alpha_s = -0.0041 \pm 0.0067$. Finally, the authors in Palanque-Delabrouille et al. (2015) reported a $\sim 3\sigma$ tension on n_s when we combined Planck 2015 and the DR9 BOSS Lyα measurement, which yielded $\alpha_s = -0.0117 \pm 0.0033$.

The analysis of Palanque-Delabrouille et al. (2015), however, was simplified : the effect of running on the Lyα likelihood was approximated as a change in the spectral index following the relation $n_s(k) = n_s(k_p) + \alpha_s \times \ln(k/k_p)$. Here, we performed dedicated simulations accounting for the full effect of running on the primordial spectrum, and added α_s to the list of parameters in the Taylor expansion of the flux power spectrum and in the Lyα likelihood. The result of our combined fit of CMB and Lyα data for the ΛCDMν+α_s cosmology are presented in Tab. 7.8. The value of the spectral index reported here is defined at the pivot scale $k_* = 0.05$ Mpc^{-1}.

Allowing a running of n_s improves the global fit to P18+lens+BAO+Lyα by $\Delta\chi^2 \sim 8.2$ compared to the plain ΛCDMν model. In a model with non-zero running, the Lyα data are compatible with significantly larger values of Ω_m than in the ΛCDMν model. With a large negative running, the effective slope of the spectrum on Lyα scales can be small (corresponding to a small effective n_s(Lyα)), and thus Ω_m can be large. We obtain a detection of running at the $\sim 3\sigma$ level. With both Pl8+Lyα and Pl8+Lyα+lensing+BAO data sets, we find $\alpha_s = -0.010 \pm 0.004$, in agreement with the previous measurement of Palanque-Delabrouille et al. (2015).

7.2.4.2 Physical interpretation

The combination of Lyα and CMB data presents several advantages. CMB data alone is more sensitive to $\sum m_\nu$ than Lyα data alone, through CMB lensing, the integrated Sachs Wolfe effect, and the measurement of the angular diameter distance to recombination (Lesgourgues and Pastor, 2006; Lesgourgues et al., 2013; Tanabashi et al., 2018). Combining the data sets helps in breaking degeneracies between cosmological parameters, such as between $\sum m_\nu$ and σ_8 (or A_s), n_s and Ω_m. This contributes to further tightening the constraint on $\sum m_\nu$.

The constraint on $\sum m_\nu$ coming from Lyα alone (with the H_0 prior) are included in Tab. 7.3, and those from P18 or P18+lens+BAO in Tab. 7.5. The joint bounds are presented in Tab. 7.6 for the ΛCDMν model, Tab. 7.7 for the case with two independent spectral indices, and Tab. 7.8 for the ΛCDMν+α_s case.

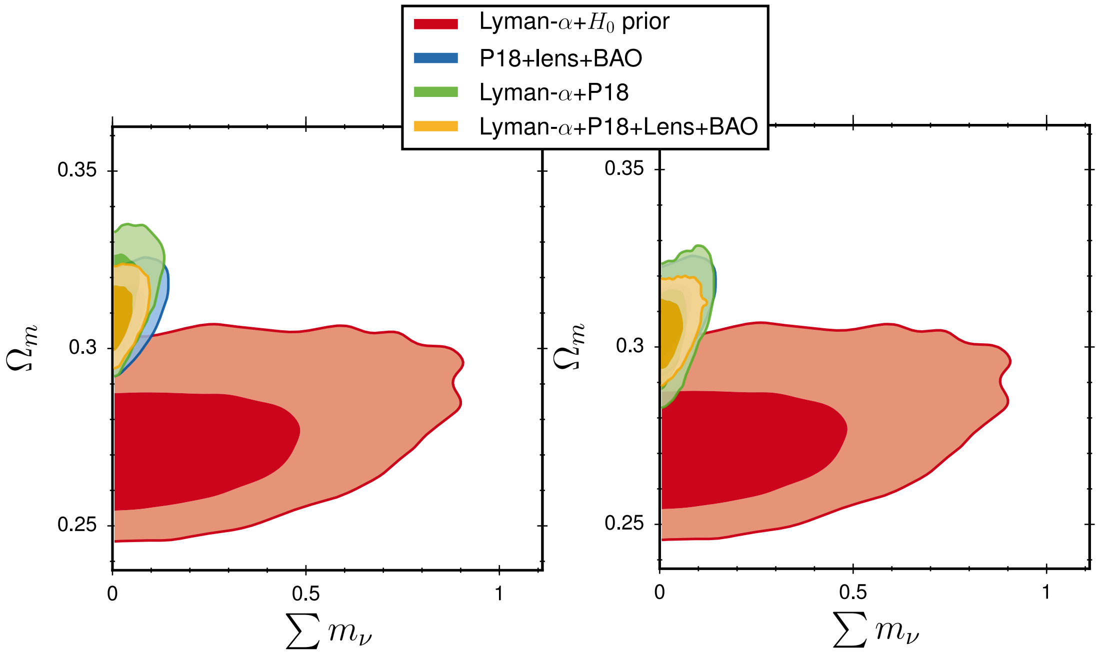

FIGURE 7.9 – *2D isocontours in the sub-space* $\{\sum m_\nu,\ \Omega_m\}$ *assuming various cosmological models and combinations of CMB, BAO and Ly*α *data. We show the* 68.3%(1σ) *and* 95.4%(2σ) *limits.* **Left :** ΛCDMν *model, showing a mild tension.* **Right :** ΛCDMν *model with two independent tilts in the Planck and Ly*α *likelihoods.*

The joint Lyα+P18+lens+BAO on $\sum m_\nu$ are loosened by a moderate amount (16%) when switching from the ΛCDMν fit (bound of 0.89 meV at 95%CL) to the fit with two free spectral indices (bound of 0.104 meV). To understand this, we can look at 2D contours in the $(\Omega_m, \sum m_\nu)$ space for these two models, shown in Fig. 7.9. The combined bounds are strongly influenced by the fact that Lyα data remove a degeneracy between Ω_m and $\sum m_\nu$ in the CMB+BAO data. Thus the neutrino mass bounds must be discussed together with the mild $\sim 2\sigma$ tension between the values of Ω_m preferred by the two data sets. Since the tension is the strongest in the ΛCDMν model, the neutrino mass is the most constrained in this case. In the model with two spectral indices, the Lyα data accommodate slightly larger values of Ω_m, because they are partly compensated by lower values of n_s at the level of the flux power spectrum. Thus the tension is relaxed and the neutrino mass bound gets a bit looser. The same trend is true in the ΛCDMν+α_s case, but the neutrino mass bounds remain slightly tighter than in the case with the two n_s values.

It is remarkable that all the neutrino mass bounds presented here fall within 20% of each other, despite the very different assumptions on the scale dependence of the spectral index. Our results for $\sum m_\nu$ are thus found to be robust against various assumptions on the possible origin of the mild tension between Lyα and Planck data when assuming a ΛCDMν cosmology. As a final result for neutrino masses, we choose to highlight the bounds coming from the most conservative of our two analyses based on a physical model, namely, the ΛCDMν+α_s analysis. We obtain $\sum m_\nu < 0.11$eV (resp. < 0.09eV) at the 95% confidence level for Lyα+P18 (resp. Lyα+P18+lens. +BAO) data, both from the frequentist or the Bayesian approach.

Given the results on neutrino oscillations (e.g. de Salas et al. (2018) for a review), these results put marginal tension on the inverted neutrino mass hierarchy scenario, which predicts a lower bound of 99.5 meV for $\sum m_\nu$. Note however that an analysis adopting the oscillation prior $\sum m_\nu > 0.05$ meV would return a looser bound than our analysis assuming $\sum m_\nu > 0$. Thus, inverted hierarchy cannot be considered as disfavored at the 2σ level by our data set.

7.2.5 Results on WDM models

To constrain the mass of a WDM thermal relic, we use a Lyα data set consisting of the eBOSS DR14 data combined with the XQ-100 data from Yèche et al. (2017). The use of high-resolution data to constrain WDM has recently been subject to debate Garzilli et al. (2015, 2019). On the one hand, they are expected to be more sensitive to WDM as they better probe the scale range impacted by the WDM power suppression, as illustrated in Fig. 7.10. On the other hand, this cut-off is partially degenerate with a similar effect caused instead by a warm IGM.

In addition of the *G_BASE* simulations we use in addition the simulations from the *G_WDM* grid to allow varying the mass of thermal relics. As explained in Chapter 6, these WDM simulations depend on the nature of WDM only through the input linear power spectrum. Also, I recall the scaling relation between non-resonantly produced sterile neutrinos and thermal relics,

$$\frac{m_{\nu_s}}{3.90\ \mathrm{keV}} = \left(\frac{m_x}{\mathrm{keV}}\right)^{1.294} \left(\frac{0.25 \times 0.7^2}{\omega_{DM}}\right)^{1/3} . \tag{7.16}$$

which is related to equation (1.91) by using $\kappa = 3.90$ keV and $\mu = 1.294$. Thus, these simulations can also be used to derive bounds on non-resonantly produced sterile neutrinos. Other classes, such as resonantly-produced sterile neutrinos, are of particular relevance. One cannot, however, constrain their mass through a simple rescaling of m_x, this would require an additional set of simulations.The same model of the IGM thermal history as in the previous sections, allowing, in addition, variations of z_{reio} as explained in Sec. 7.2.1.2.

The eBOSS DR14 + XQ-100 Lyα data constrain the WDM mass to $m_x > 8.6$ keV (95% CL). Because non-linear structure growth tends to erase the power suppression caused by the free-streaming of relativistic particles, the sensitivity to m_x is more prominent at high redshift. Thus, the high-redshift data are the ones with the largest constraining power, despite having the largest statistical uncertainties. This feature is confirmed by the loosening of the bound as we restrict the data to lower redshifts. We obtain $m_x > 5.3\,\mathrm{keV}$ for $z < 4.5$, and $m_x > 3.9\,\mathrm{keV}$ for $z < 4.1$, both at 95% CL. In Tab. 7.9, we give the best-fit values for thermal relics warm dark matter and the associated bounds on non-resonantly produced neutrinos using the rescaling relation for different redshift cuts of the SDSS Lyα data set. This trend is also clearly visible in Fig. 7.11, where we show the χ^2 profiles for the three configurations. The profile widens (hence the constraint loosens) as we go from all redshifts (dark blue) to $z < 4.5$ (blue) then to $z < 4.1$ (light-blue).

The gain provided by the highest-redshift $z = 4.6$ bin, however, is two-fold. It is due in part because the $z = 4.6$ power spectrum probes structure growth in a more linear regime, but also because the minimum of the fit then occurs for $1\,\mathrm{keV}/m_x < 0$. The Feldman-Cousins prescription Feldman and Cousins (1998) allows us to derive the limit in such a case, by computing the $\Delta\chi^2$ with respect to the χ^2 at the limit of the physical domain, i.e. where $1\,\mathrm{keV}/m_x = 0$. This method, however, leads to an artificially strong limit, since the χ^2 profile in the physical ($m_x > 0$) region is a very steep function of $1\,\mathrm{keV}/m_x$. To be conservative, we hence consider as our main result the bound obtained in the eBOSS DR14 ($z < 4.5$) + XQ-100 configuration : $m_x > 5.3\,\mathrm{keV}$ (95% CL). We highlight the fact that removing the highest $z = 4.6$ redshift bin

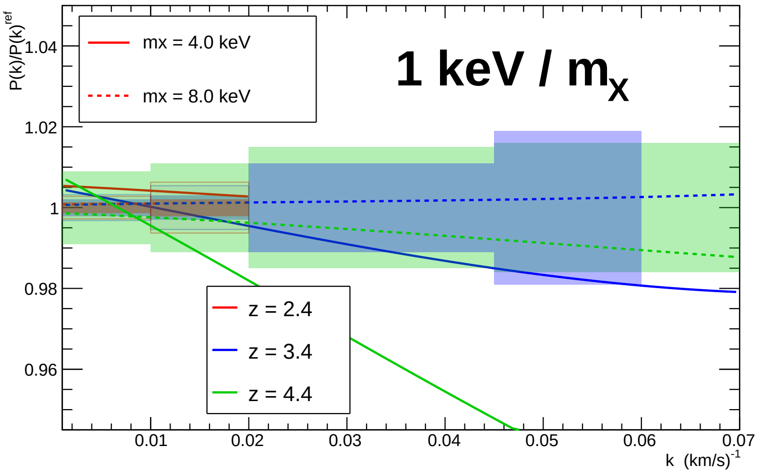

FIGURE 7.10 – *Sensitivity of the eBOSS and XQ-100 data to WDM. The curves show the predicted deviation of the Lyα power spectrum for a WDM particle mass of 4.0 keV (solid) or 8.0 keV (dashed) relative to* ΛCDM, *in three redshift bins, when all other cosmological, astrophysical and nuisance parameters are kept fixed. The boxes illustrate the data uncertainty using the same code as for the curves. The z=2.4 bin only refers to eBOSS data and is bounded to* $k < 0.02\,\mathrm{s{\cdot}km^{-1}}$. *Shaded boxes refer to statistical errors only, while larger clear boxes bounded by colored lines refer to statistical plus systematic errors. Systematics only significantly impact the total uncertainty for* $k < 0.02\,\mathrm{s \cdot km^{-1}}$ *(eBOSS data) and at small redshift* ($z < 3.4$ *typically).*

	m_x (keV , 95% CL)	m_{ν_s} (keV , 95% CL)
Lyα+XQ-100	8.6	63
Lyα(z<4.5)+XQ-100	5.3	34
Lyα(z<4.1)+XQ-100	3.9	22

TABLE 7.9 – *Constraints on mass bounds for thermal relics and non-resonantly produced sterile neutrinos (68.3% confidence level) in a* ΛWDM *model by combining Lyα and XQ-100 data with different redshift cuts for Lyα SDSS data.*

is sufficient to ensure that the best-fit minimum is in the physical region. Further data restriction does not alter its location. In Tab. 7.10, we give the best-fit values of some of the main parameters.

These new bounds are significantly tighter than the 4.2 keV lower limit derived with BOSS DR9 ($z < 4.5$) + XQ-100 by Yèche et al. (2017) (red curve in Fig. 7.11). The improvement mostly comes from the improved statistical power of the eBOSS DR14 data compared to BOSS DR9, as is made clear by the comparison of the the limits obtained with the same $z < 4.5$ redshift range.

As our most robust bound on WDM, we therefore take the eBOSS ($z < 4.5$) + XQ-100 configuration, marginalizing over the cosmological, astrophysical and nuisance parameters described in Tab. 7.2. This leads to $m_x > 5.3\,\mathrm{keV}$ (95% CL), or equivalently to a constraint on the mass of a non-resonantly produced sterile neutrino $m_{\nu_s} > 34\,\mathrm{keV}$ (95% CL).

	Lyα(z<4.5)+XQ-100
T_0 (z=3) (10^3K)	13.9 ± 1.7
γ	1.09 ± 0.13
σ_8	0.796 ± 0.020
n_s	0.954 ± 0.006
Ω_m	0.265 ± 0.008
m_x (keV , 95% CL)	> 5.3

TABLE 7.10 – *Preferred astrophysical and cosmological parameter values (68.3% confidence level) for the WDM model, for combined Lyα($z < 4.5$) and XQ-100 data.*

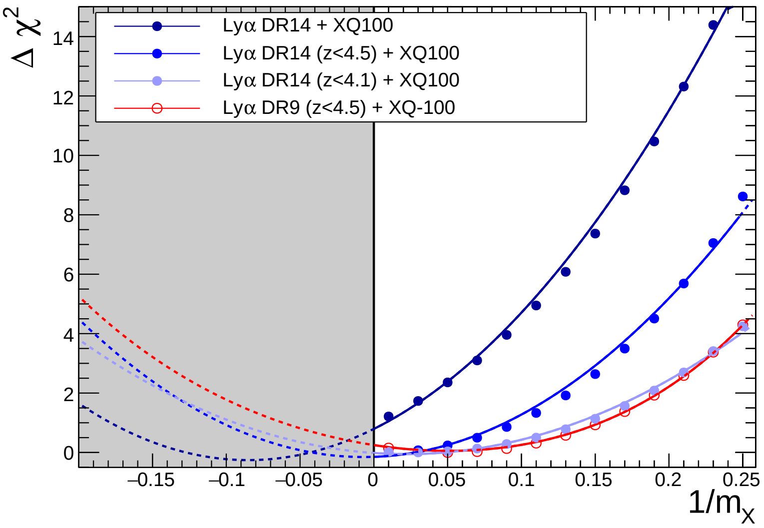

FIGURE 7.11 – $\Delta\chi^2$ *profile as a function of* $1\,\mathrm{keV}/m_x$ *for the four configurations : Lyα+ XQ-100 (dark blue), Lyα($z<4.5$) + XQ-100 (blue), Lyα($z<4.1$)+ XQ-100 (light blue), DR9 Lyα($z<4.5$) + XQ-100 (red with open circles). Each point shows the* $\Delta\chi^2$ *obtained from a profiling method, i.e. after minimization over all other free parameters. The curves are the result of a parabolic fit to the points, extrapolated into the negative region.*

8

Conclusion and prospects

Contents

8.1 General conclusions of this thesis

The contemporary ΛCDM cosmological model is an impressively successful framework for predicting and explaining many independent observations, from the CMB to large-scale structures. However, cosmology is far from having reached its final goal since the concordance model still faces many challenges. First, it remains a phenomenological model since the actual physics of dark matter and dark energy are still not understood. In addition, we have now entered the era of precision cosmology meaning that the precision of measurement of cosmological parameters have reached the percent level. Thus, even if all data converge toward the ΛCDM model, some independent probes now exibit tension at the 5σ level. These divergences could be the results of unidentified biases in data analysis or could reflect the need of new physics beyond the standard model. Even if these deviations appear to be small, they are significant enough to have the power to discriminate between different cosmological models.

In this context, my PhD work tackles some of these fundamental issues. I provide constraints on the sum of the masses of active neutrinos, which is a key element to go beyond the Sandard Model of Particle Physics, necessary to understand the neutrino oscillation mechanism that proves its limits. In addition, I investigate the small-scale challenges of the cold dark matter scenario by determining the plausibility of a warm dark matter model. Investigating these fundamental issues at CEA Saclay has been a truly fulfilling and fruitful experience.

To provide these constraints, I measured the impacts of a massive neutrino model and a warm dark matter model on the formation of structures in the universe. The most efficient cosmological observable to probe their effects is the 1D power spectrum of the Lyα forest that ideally probes small scales, which are affected by the free streaming of these relativistic particles, on a range of redshifts where the evolution of fluctuations are not erased by non linearities yet. Chapter 2 presents the physics of the Lyα forest, which is an imprint of the neutral atomic hydrogen in the IGM. Extremely luminous quasars act as background beacons to study in absorption neutral clouds of hydrogen along lines of sight, which constitute the Lyα forest.

To measure the 1D power spectrum of the Lyα forest I used quasar spectra from the BOSS and eBOSS programs of the spectroscopic SDSS-III and SDSS-IV surveys, fully described in Chapter 3. Chapter 4 presents the measurements of the 1D power spectra of the Lyα forest using 43,751 high-quality quasar spectra with absorbing redshift $2.1 \leq z_{\mathrm{Ly}\alpha} \leq 4.7$. Compared to previous measurements, this study benefits from more than a factor three improvement on the number of quasars. The required tolerance of the systematic budget must increase accordingly to the statistical gain. Therefore, to fully take advantage of this unprecedented statistical power, I performed a careful investigation of all the observational systematic errors and their sources in order to provide robust constraints on the cosmological parameters. Instrumental systematics (noise modeling, spectral resolution...) dominate the small scales, which are sensitive to $\sum m_\nu$, and the hypothetical warm dark matter component. Method systematics dominate the large scales, which can probe the amplitude of density fluctuations σ_8 or the scalar spectral index n_s.

In particular, the quasar-continuum modeling and the effects of high residual absorbers represent twice the statistical uncertainties on large scales.

The increased statistical power is such that systematics are now becoming the dominant source of uncertainties. Controling observational and theoretical systematics is fundamental to not hamper the physical interpretation of the data. Because the Lyα forest probe scales in the midly non-linear regime and because we observe light flux from quasars, we cannot estimate theoretical predictions using analytical recipes. Hence, we have to rely on cosmological hydrodynamical simulations which are reviewed in Chapter 5. In particular, I identify the major limitations and issues of numerical simulations in the context of the Lyα forest modeling. If it was initially thought that gravitational instabilities and hydrodynamics alone were the only significant processes to accurately model the Lyα forest, we now know that it results from the complex interplay of large-scale structure evolution driven by the dark-matter density fluctuations, and small-scale baryonic physics. Indeed IGM gas might be consequently affected by galaxy evolution because of gas cooling, star formation, cosmic reionization the feedbacks from supernovae and AGN that inject energy in the ambient medium. Including these processes in simulations requires a huge dynamic range, which is computationally expensive and hardly compatible with a thorough coverage of the cosmological parameter space. Furthermore their implementation in simulations rely on arbitray parameters calibrated on astrophysical observations, such as the stellar mass of galaxies or their star formation rate, leading to discrepancies between different hydrodynamical simulations.

Chapter 6 describes the improvement of the grid of hydrodynamical simulations covering the parameter space of active neutrinos and thermal warm dark matter that lacks baryonic feedbacks. I constrained the impact of AGN feedback on the Lyα power spectra using a series of 8 hydro-cosmological simulations that I performed with the grid-based Eulerian code RAMSES using 10 million hours on 4096 to 8192 computer cores. I start from the Horizon-AGN simulation and vary the sub-grid parameters for AGN feeding and feedback to cover the whole plausible range of feedback models, according to the resulting galaxy properties, rather than relying on a single specific implementation. I show that AGNs globally suppress the Lyα power at all scales. On large scales, the energy injection and ionization dominate over the supply of gas mass from AGN-driven galactic winds, thus suppressing power. On small scales, faster cooling of denser gas mitigates the suppression. This effect increases with decreasing redshift. I provide lower and upper limits of this signature at nine redshifts between $z = 4.25$ and $z = 2$, making it possible to account for it at post-processing stage. I also show that ignoring AGN feedback in Lyα analysis leads to strong biases with 2% shift on σ_8 and 1% shift on n_s , which represents twice the standard deviation of the current constraints on n_s.

Chapter 7 combines the observation measurements and theoretical predictions from hydrodynamical simulations to constrain cosmology. First, I inverted the signal from many cosmological observables, including CMB, galaxy clustering, weak lensing along with the Lyα forest, to estimate the total matter power spectrum at $z = 0$ spanning several orders of magnitude in physical scale and in cosmic history. I developed a new lower-noise method for performing the inference from the Lyα power spectra, which significantly reduces the uncertainties. These results thus highlight the good agreement of the ΛCDM model with observational data issued by independent experiments, covering a large range of cosmic times and cosmic scales. Then, I directly used the Lyα signal to constrain cosmological parameters. A mild tension is found between the values of Ω_m preferred by the Lyα and the CMB data. Interestingly, the Lyα best-fit cosmological parameters are in very good agreement with current weak lensing constraints on (Ω_m, σ_8). Lyα

and weak lensing are two late-time probes of a similar range of scales, and they show a comparable level of tension with Planck ΛCDM model at the $2-3\ \sigma$ level. Because Ω_m and n_s have a similar impact on the Lyα flux power spectrum, the small tension on Ω_m is likely to have the same origin as the mild tension on n_s previously observed in SDSS Lyα analysis. However, we showed that it could be reduced by considering different scalar indices on CMB and Lyα scales, such as produced by a running of the scalar index. We find a mild preference for a non-zero running of n_s at the level of about 3σ : $\mathrm{d}n_s/\mathrm{d}\ln k \sim -0.010 \pm 0.004$. It illustrates the small disagreement in the slopes of the power spectrum measured independently by BOSS/eBOSS and Planck. The free-streaming of massive neutrinos causes a step-like suppression in the power spectrum that is ideally probed by comparing the large-scale CMB to the small scale Lyα power spectra. The constraint on $\sum m_\nu$ thus comes from the measurement of σ_8 in Lyα data, and from the correlation between σ_8 and $\sum m_\nu$ provided by CMB. Combining BOSS and eBOSS Lyα with Planck CMB data, we find an upper bound on $\sum m_\nu$ of 0.10 eV (95% CL) for a ΛCDM model, which only loosens to 0.11 eV when allowing for running. When further including CMB lensing and BAO, the limit tightens to $\sum m_\nu < 0.09$ eV, whether or not running is allowed. These limits improve over those previously published and are the best constraints to-date. They tend to favor the normal hierarchy neutrino mass scenario. WDM affects clustering compared to the CDM scenario by suppressing all power below a scale determined by the particle mass. Thanks to the much improved statistics, the BOSS and eBOSS data allow us to improve the limit on WDM compared to previous publications. Using a combination of eBOSS ($z < 4.5$) + XQ-100 Lyα data, we constrain the mass m_x of a thermal relic to $m_x > 5.3$ keV (95% CL). It translates to a constraint on the mass m_{ν_s} of a non-resonantly-produced sterile neutrino of $m_{\nu_s} > 34$ keV (95% CL).

8.2 Future prospects

As every cosmological constraints, these results are limited by both data analysis and theoretical modeling, which require a rigorous control of systematics. Thus, there are two main roads for improvements.

First, new-generation spectroscopic data from DESI, WEAVE, or 4MOST will soon be available. With a three to four-fold increase in quasar number density, it will be possible to further tighten the selection of the quasar spectra and reduce the contamination from systematic biases. With DESI, the factor of two gain in resolution and the reduced noise will improve the measurement on small scales relevant for $\sum m_\nu$ and m_x allowing to go to scales as small as 0.5 $\mathrm{Mpc}\cdot\mathrm{h}^{-1}$, when SDSS could reach 1 $\mathrm{Mpc}\cdot\mathrm{h}^{-1}$. The extension of the DESI quasar selection to higher redshift is highly relevant to WDM. With a projected resolution of 0.020 eV, DESI will make a precision measurement of the sum of the neutrino masses independent of hierarchy and therefore determine the absolute mass scale for neutrinos, a measurement that is otherwhise very challenging. Furthermore, if the masses were minimal and the hierarchy normal, DESI would be able to exclude the inverted hierarchy at 2σ. Also, DESI will probe the effective number of neutrino species $N_{\nu,\mathrm{eff}}$ which parametrizes the energy density attributed to any non-electromagnetically interacting relativistic species. The detection of any discrepancy from the expected value, $N_{\nu,\mathrm{eff}} = 3.04$, would be a truly major result, which could potentially indicate the existence of sterile neutrinos, among other possibilities. Forecasts predict measurement with a $\sim$ 10% precision or better, providing strong constraints on the alternative models.

On the modeling side, the increase of the supercomputing power is well in line with instrumental development and will allow to gain more physical insights in sub-resolution recipes, which are the main challenges in cosmological hydrodynamical simulations. Indeed, we expect the first exascale facility at the Très Grand Centre de Calcul by 2023. On the short term, bayronic processes impacting the Lyα forest at level similar to that of current data uncertainties can be properly estimated. In particular, during the last month of my thesis, I performed the EHnoAGNnoSN simulation to estimate the impact of supernovae winds on the Lyα forest power spectra. Since the analytical correction currently used comes from the OWLS set of simulations that suffers from feedbacks significantly stronger than observations, it potentially over-estimates the impact on the Lyα forest similarly to what I found for AGN feedback. Indeed, these supernovae analyses show that while the authors in Viel et al. (2013) finds a 5% correction to apply to the $P_{\mathrm{Ly}\alpha}$ we find that SN feedback has barely no impact, or at least well below the 1% level. The most impacted cosmological parameter from the ΛCDM model is n_s, which is found to be 0.950 ± 0.005 and 0.943 ± 0.005 using analytical corrections from Viel et al. (2013) or the ones from the set of EH simulations respectively. It shows a 1σ discrepancies between the two inferences, which might appear negligible for now but it is important to keep in mind that it will strongly increase with DESI data when observational systematics will decrease. This is the subject of a paper in preparation.

On the longer term, we will be able to include additional physical processes that we will not be able to overlook anymore with respect to observational uncertainties. In particular, numerical simulations will probably need to include radiative transfer to constrain the thermal state and neutral fraction of the IGM, which will allow to account for inhomogeneous reionization. This will be particularly useful for warm dark matter constraints since the small-scale cut-off in the matter power spectrum can also be attributed to the instantaneous temperature of the IGM that suppresses power on small scales via the Doppler broadening and to the pressure support that smooths the distribution of baryons compared to dark matter fluctuations and suppresses power below the filtering scale that depends on the entire thermal history.

Also, if hydrodynamical simulations are very efficient to study the $P_{\mathrm{Ly}\alpha}$, the EH simulation shows that we are close to the limit of what is possible today in terms of compromise between volume and resolution. As such, we still have to rely on analytical methods for the Lyα correlation function that can put strong constraints on dark energy models and modified gravity theory since it is sensitive to the Baryon Accoustic Oscillation (BAO) signal and the linear growth rate of structures through redshift space distorsions. Indeed, because it exploits information from different line of sights it probes a volume about ten times larger than studies of the 1D power spectrum, but it still requires to resolve the Jeans length increasing by an order of magnitude the dynamical range. On the one hand, Lyα-BAO analyses, which use information on large linear scales of about ~ 100 Mpc, construct theoretical predictions by means of analytical recipes whose robustness can be assesed against large-volume dark matter only simulations. However, these simulations completely lack the hydrodynamical aspect of the BAO signal and might bias our intepretation of the data. On the other hand, Lyα-RSD analyses use information on smaller scales, of about ~ 20 Mpc, while conserving the very large volume. If analytical theoretical models exist, their robustness must be assessed using hydrodynamical simulations given that the signal reaches midly non-linear scales where the Lyα gas distribution displays even stronger discrepancies with the dark matter distribution. So far, such tests were performed on 120 $\mathrm{Mpc}\cdot\mathrm{h}^{-1}$ boxes, but the volume was shown to be too small so that the RSD signal was dominatd by cosmic variance. Therefore, few Lyα-RSD constraints exist because of a lack of confidence in theoretical predictions that are not robust enough to be confronted with observations. In this context, I recently submitted a proposal as co-PI for 200,000 GPU hours on the new TGCC/Jean-Zay partition for the ACCELL2 project (L'expansion ACCELeré de l'univers sombre mise en lumière par le calcul ACCELeré sur GPU). We aim at producing a hydrodynamical large-volume grid, from 150 to 600 $\mathrm{Mpc} \cdot \mathrm{h}^{-1}$ resolving the IGM Jeans length to at least have $\Delta x_{\mathrm{min}} = 100\ \mathrm{kpc} \cdot \mathrm{h}^{-1}$, in order to investigate hydrodynamical impacts on Lyα- BAO and RSD signals and to provide robust tests of analytical methods. Such a project could significantly contribute to provide tests of General Relativity and dark energy models on a range of redshift unreachable by other tracers such as galaxy or quasars.

To conclude, thanks to our increasing capacity to develop new powerful technologies, on both observation and theoretical sides, the cosmic gold mine role of the Lyα forest is going to strengthen over the next years. There is no doubt that the parallel development of independent cosmological probes will provide strong tests of the ΛCDM model, on the fundamental nature of dark matter, dark energy, on the validity of General Relativity, on the physics of the early universe or on the existence of still unknown components such as sterile neutrinos. I believe this is a very exciting time to study cosmology, and I look to contributing further on unveiling the mysterious dark universe.

A

Formation of compact galaxies in the Extreme-Horizon simulation

Early-type galaxies (ETGs) at redshift $z > 1.5$ are much more compact than nearby ones (Daddi et al., 2005). At stellar masses about $10^{11}\,\mathrm{M}_\odot$, they typically have half-mass radii of 0.7–3 kpc, about three times smaller than nearby ellipticals with similar masses (van der Wel et al., 2014). Compact radii come along with steep luminosity profiles and high Sersic indices (van Dokkum and Brammer, 2010; Carollo et al., 2013). Star-forming galaxies (SFGs) also decrease in size with increasing redshift (e.g., Kriek et al., 2009; Dutton et al., 2011). Besides, the CANDELS survey has discovered a population of very compact SFGs at z$\sim$2 : the so-called "blue nuggets" (Barro et al., 2013; Williams et al., 2014) have stellar masses of $10^{10-11}\,\mathrm{M}_\odot$ with unusually small effective radii around 2 kpc and sometimes even below 1 kpc. Compact SFGs have high comoving densities, about $10^{-4}\,\mathrm{Mpc}^{-3}$ for stellar masses above $10^{10}\ \mathrm{M}_\odot$, and $10^{-5}\,\mathrm{Mpc}^{-3}$ above $10^{11}\,\mathrm{M}_\odot$ (Wang et al., 2019). In addition, SFGs at $z \simeq 2$ often have very compact gas and star formation distributions (Elbaz et al., 2018).

Many processes have been proposed to explain the formation of compact galaxies, ranging from early formation in a compact Universe (Lilly and Carollo, 2016) to the compaction of initially-extended galaxies (Zolotov et al., 2015) through processes that may include galaxy mergers, disk instabilities (Bournaud et al., 2007; Dekel and Burkert, 2014), triaxial haloes (Tomassetti et al., 2016), accretion of counter-rotating gas (Danovich et al., 2015) or gas return from a low-angular momentum fountain (Elmegreen et al., 2014).

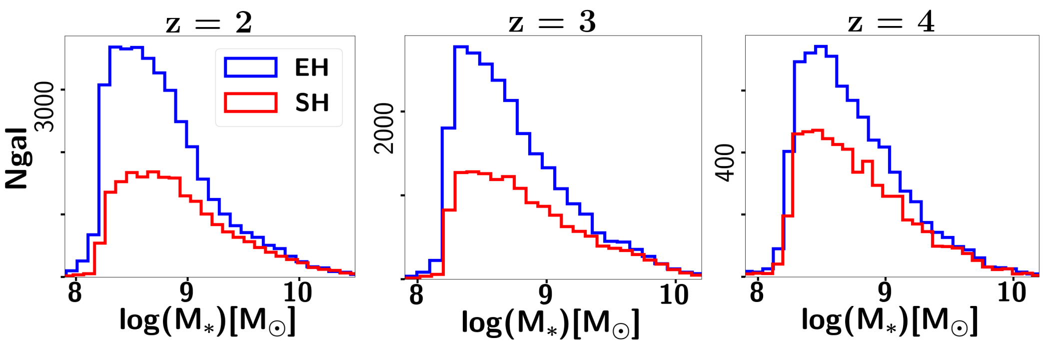

Figure A.1 – *Number of galaxies per mass bin in EH and SH at $z = 2$, 3 and 4.*

The Extreme-Horizon (EH) cosmological simulation, presented in Sec. 6.3.1, models galaxy-formation processes with the same approach as Horizon-AGN (HAGN, Dubois et al., 2016, hereafter D14) and a substantially increased resolution in the intergalactic and circumgalactic medium (IGM and CGM). The properties of massive galaxies in EH and the origin of their compactness are studied in Sec. A.1.2 and Sec. A.1.3. This work has been recently submitted with the ensemble of the Extreme-Horizon participant (Chabanier et al., 2020).

A.1 Galaxy compaction in EH

A.1.1 Galaxies in the EH simulation

We detect galaxies with more than 50 stellar particles (about $10^8\,\mathrm{M_\odot}$) using AdaptaHOP (Aubert et al., 2004). 37,698 galaxies are detected in EH at $z \sim 2$ and 20,314 in SH, with stellar mass functions at various redshifts shown in Fig. A.1. While the mass functions above $10^{10}\,\mathrm{M_\odot}$ are quite similar in both simulations, EH forms twice as many galaxies as SH with stellar $\mathrm{M_*} \leq 5 \times 10^9\,\mathrm{M_\odot}$. We rule out any detection bias since stellar particles have similar masses in EH and SH (new stars form at the maximal resolution level in each simulation), and attribute this difference to the increased resolution in low-density regions. Fitting the $z = 2$ mass function with a power-law of the form $\Phi(M_*) \propto M_*^\beta$ in the $10^9 \leq \log(M_*/M_\odot) \leq 10^{9.5}$ range yields $\beta = -0.68$ for EH and -0.34 for SH. Observations indicate a slope $-1.0 \leq \beta \leq -0.5$ in this mass range (Santini et al., 2012; Tomczak et al., 2014), showing that low-mass galaxy formation is substantially under-resolved or delayed in SH.

We build samples of galaxies with $M_* \geq 5 \times 10^{10}\,\mathrm{M_\odot}$. On-going major mergers identified through the presence of a companion with more than 20% of the stellar mass within 20 kpc and/or a double nucleus, are rejected, yielding a sample of massive galaxies displayed in Fig. A.2 and in Fig. **??**.

We then study the mass distribution of the selected galaxies, taking into account non-sphericity. Stellar density maps are computed with a 500 pc pixel size. Pixels below $50\,\mathrm{M_\odot} \cdot \mathrm{pc}^{-2}$, typically corresponding to a surface brightness $\mu_i \geq 28\,\mathrm{mag} \cdot \mathrm{arcsec}^{-2}$, are blanked out. Ellipse-fitting of iso-density contours is performed using the technique from Krajnović et al. (2006) taking into account the 500 pc pixel size and the ~ 1 kpc PSF. Satellite galaxies are removed as follows : the circular region centered on the luminosity peak of the companion and extending up to the saddle of the luminosity profile between the main galaxy and the companion is ignored in the ellipse-fitting procedure, and replaced with the density profile modeled on other regions. Satellites with a mass below 5% of the main galaxy are ignored to avoid removing sub-structures

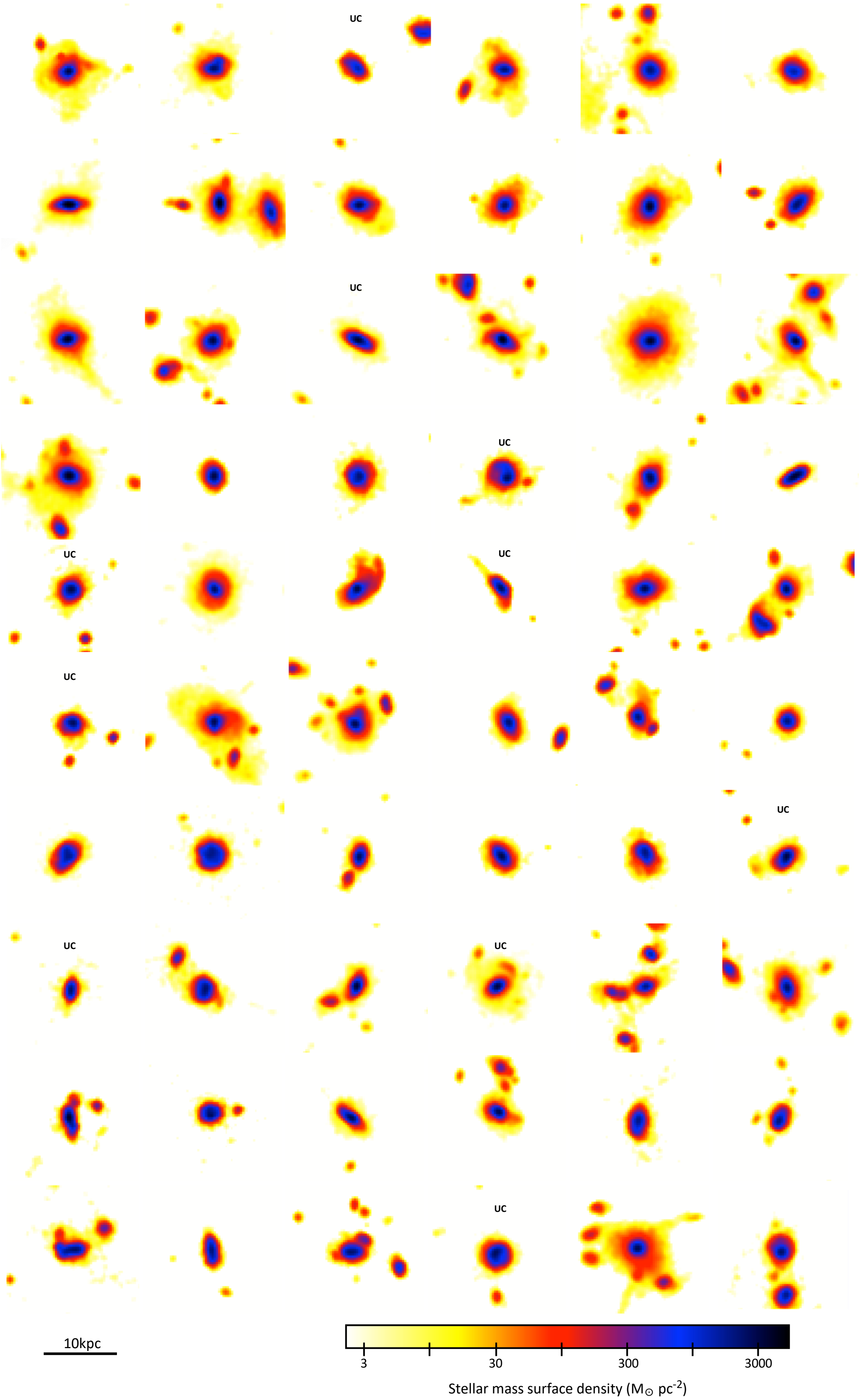

FIGURE A.2 – *Stellar mass distribution of massive galaxies in the EH simulation. UC galaxies are flagged.*

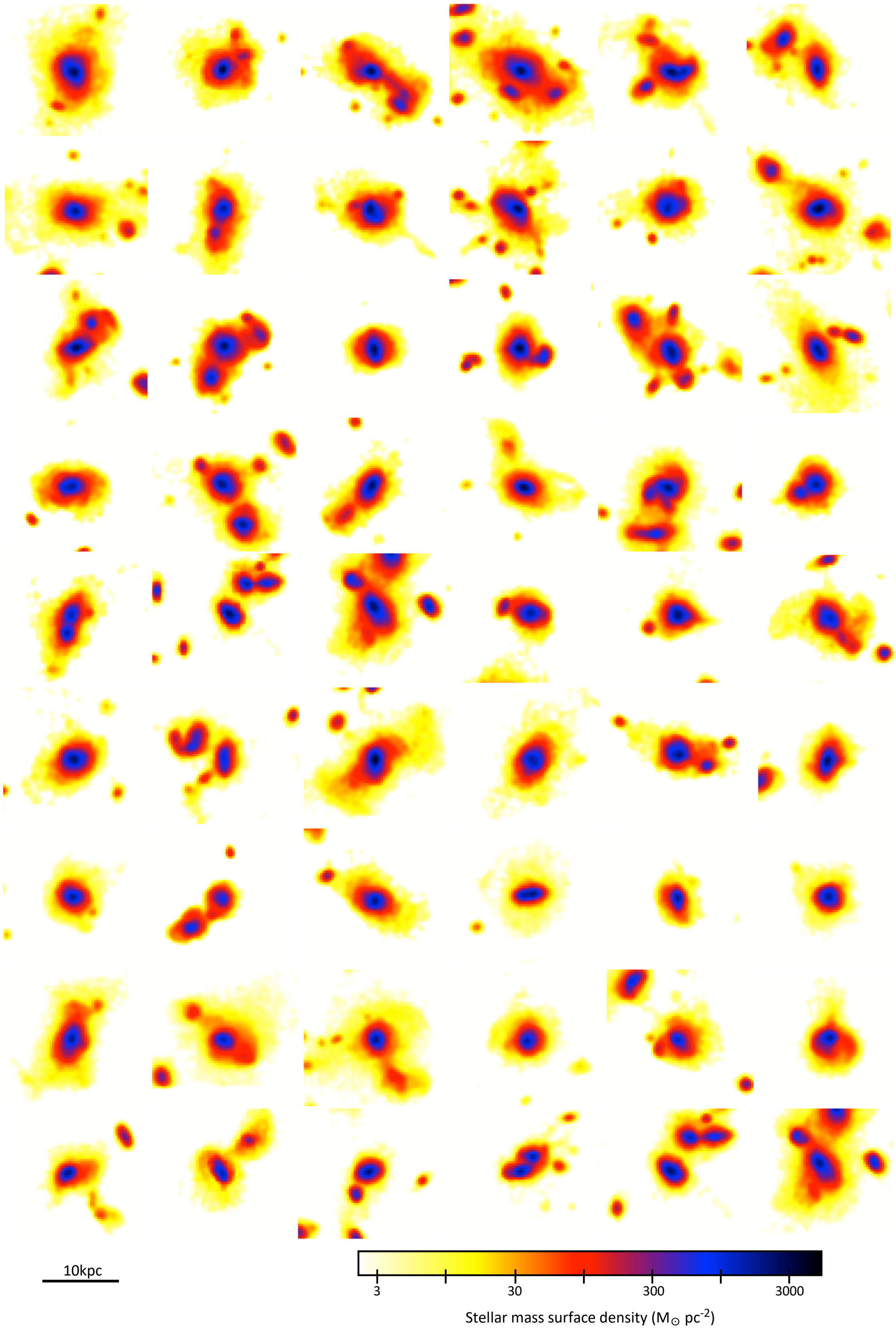

FIGURE A.3 – *Stellar mass distribution of massive galaxies in the SH simulation. Galaxies are not meant to be individually matched to SH galaxy as independent samples were built in EH and SH.*

Galaxy samples	Mean difference in $\mathcal{L}_{in}$
EH vs. SH	13% lower
EH-NUC vs. SH	10% lower
EH vs. SH at $M_* < 10^{11}\,\mathrm{M}_\odot$	12% lower
EH-UC vs. EH-NUC at $M_* < 10^{11}\,\mathrm{M}_\odot$	3% lower

Table A.1 – *Mean difference in the specific angular momentum of inflowing gas $\mathcal{L}_{in}$ between several samples of massive galaxies.*

of the main galaxy. Three perpendicular projections are analyzed for each galaxy, and the median results are kept for both the stellar mass M_* and the half-mass radius R_e, the latter being defined as the semi-major axis of the isophote-fitting ellipse containing 50% of the stellar mass. The removal of satellite galaxies and low-density outskirts yields final stellar masses slightly below the initial estimates, down to $M_* \simeq 3 - 4\times 10^{10} \cdot M_\odot$.

Stellar masses, radii and compactness distributions for EH and SH samples are shown in Fig 6.17. We see that massive galaxies in EH are globally more compact than in SH, and EH contains a population of ultra-compact (UC) outliers. The larger sizes in SH do not just correspond to extended stellar haloes : the difference remains when we vary the surface density threshold in mock images, and Sersic indices are on average similar in EH and SH. The size difference is not expected to arise from internal processes such as instabilities and/or feedback, as galactic scales and feedback are treated with the very same resolution in EH and SH. Two key differences could contribute : EH models gas flows in the CGM at a much higher resolution, and low-mass galaxies are under-resolved in SH.

A.1.2 Diffuse accretion and angular momentum supply

A substantial part of the angular momentum of galaxies is supplied by cold gas inflows (Ocvirk et al., 2008; Pichon et al., 2011; Danovich et al., 2015; Tillson et al., 2015) which are better resolved in EH. Higher resolution could also better probe metal mixing in the IGM and subsequent cooling (Scannapieco et al., 2006). To probe these potential effects, we focus on inflowing gas in the vicinity of massive galaxies using the following criteria, which typically select inflowing gas according to other simulations (e.g., Goerdt et al., 2015) :

- a galactocentric radius between $3\,R_e$ and 50 kpc,
- a density below $0.1\,\mathrm{cm}^{-3}$ to exclude satellites,
- a velocity vector pointing inwards w.r.t. the galaxy center,
- a temperature below $10^{5.5}$ K.

For each resolution element following these criteria, we compute the gas mass m and angular momentum l w.r.t the galaxy center (in norm, $l = \|\vec{l}\|$), sum-up the total angular momentum $L = \Sigma\, l$ and mass $M = \Sigma\, m$ for inflowing gas, and compute the angular momentum of inflowing gas $\mathcal{L}_{in} = L/M$ around each galaxy. Differences in $\mathcal{L}_{in}$ for various galaxy samples are listed in Table A.1, showing that $\mathcal{L}_{in}$ around massive galaxies is substantially lower in EH than in SH, but is almost similar around EH-UC and EH-NUC galaxies. Taking the magnitude of the vector sum instead yields results in agreement to within a few percent, and leads to the same general conclusion [1].

---

[1] The differences in angular momentum of inflowing gas listed in Table A.1 are marginally larger when considering the magnitude of the vector sum.

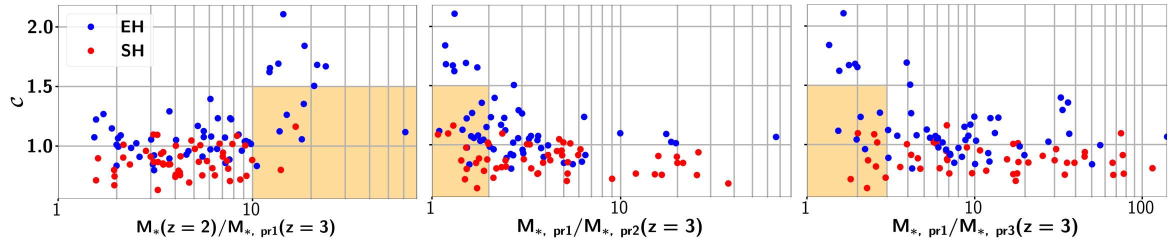

FIGURE A.4 – *Compactness $\mathcal{C}$ as a function of the ratio between the stellar mass at $z = 2$, $M_*(z = 2)$, and the mass of the main progenitor at $z = 3$ $M_{*,\ \mathrm{pr1}}(z = 3)$ (left panel) and as a function of the ratios between the mass of the three most massive progenitors (first and second, $M_{*,\ \mathrm{pr2}}$ on the middle panel and first and third, $M_{*,\ \mathrm{pr3}}$, on the right panel) for EH (blue) and SH (red) galaxies. The shaded areas define regions of galaxies that would grow through major mergers of low-mass galaxies but would end-up NUC : no SH galaxy and only one EH galaxy are in the three areas at the same time.*

We can estimate the potential impact on galaxy sizes under two extreme assumptions. On the one hand, if the circular velocity remains unchanged, dominated by a non-contracting DM halo, then galactic radii should follow $R \propto \mathcal{L}_{in}$. On the other hand, if the dark matter halo contracts in the same proportions as the baryons, the rotation velocity V and radius R follow $V^2 \propto 1/R$ at fixed mass, so that $R \propto \mathcal{L}_{in}^2$.

Hence the 10% difference in $\mathcal{L}_{in}$ between EH-NUC and SH could result in a 10–20% size difference : this can account for the smaller sizes of massive galaxies in EH compared to SH. On the other hand, the population of UC galaxies does not result from diffuse gas accretion as it could only impact sizes by a few percent compared to NUC galaxies.

Angular momentum is built up by tidal torques that only depend on very large-scale structures expected to be well resolved even in SH (Fall and Efstathiou, 1980). Yet, angular momentum can be lost when cold inflowing streams interact with hot gas haloes and outflows in the CGM. Idealized simulations of cold streams interacting with hot haloes (Mandelker et al., 2020) indicate that instabilities can decrease the velocity of cold streams by up to a few tens of percent in favorable cases, which can explain the loss of angular momentum at the EH resolution compared to SH[2].

A.1.3 Major mergers of low-mass progenitors

Another driver of compaction could be the numerous low-mass galaxies in EH that are missing in SH. We identify the progenitors of $z = 2$ UC and NUC galaxies by tracking their stellar particles, and analyze their progenitors at $z = 3$ and $z = 4$ with the same technique as our $z = 2$ sample.

Figure. A.4 shows the compactness as a function of the mass ratios between each $z = 2$ galaxy and its main $z = 3$ progenitor and between the main $z = 3$ progenitor and the second and third most massive progenitors. UC galaxies have (1) a main $z = 3$ progenitor that never exceeds 10% of the $z = 2$ mass, (2) a second and (3) third most massive progenitors almost as massive as the main progenitor, with mass ratios lower than 3 :1 (generally lower than 2 :1) for the second most massive, and generally below 4 :1 for the third most massive. This pinpoints a correlation between these parameters, showing that the formation of EH-UC galaxies involves

---

[2] Mandelker et al. suggest that 10–20 resolution elements per stream diameter are required to model such instabilities. For our typical filament diameter of 20–30 kpc at z=2–3, EH reaches such resolution in the CGM, but SH does not (Fig. 6.13).

repeated[3] major mergers between low-mass progenitors. These mergers occur rapidly between $z = 3$ and $z = 2$ with 80% of UC galaxies that assemble 90% of their stellar mass in this redshift range. Conversely, 70% of galaxies that have assembled 90% of their stellar mass between $z = 3$ and $z = 2$ end up as UC galaxies.

In contrast, EH-NUC and SH galaxies most often have one dominant progenitor undergoing only minor mergers, and very rarely meet the three criteria depicted above for UC formation at the same time. There is actually no SH galaxy and only one EH-NUC galaxy that lies in the three shaded areas in Fig. A.4 at the same time. This strengthens our argument that these specific types of accretion histories essentially always produce UC galaxies. The only exception among EH-NUC galaxies has an extended spiral disk morphology, and has the second highest total angular momentum L in inflowing gas over the whole EH sample so that accretion of diffuse gas compensates for the compacting effects of the merger history in this extreme object. It is expected from idealized simulations of repeated mergers with various mass ratios that mergers histories involving mostly major mergers with relatively similar masses produce more concentrated end-products for the same total merged mass (at least in terms of Sersic indices, Bournaud et al. 2007, Fig. 4). 45% and 47% of the stars found in EH-NUC and SH galaxies at $z = 2$ are already formed at $z = 3$, respectively, compared to only 36% for EH-UC galaxies : UC galaxies arise from low-mass progenitors and hence form their stars later on.

We also note that the distributions of progenitor masses are fairly identical for EH-NUC and SH galaxies (Fig. A.4) indicating that the smaller sizes of EH-NUC galaxies doe not result from different merger histories but rather from the modeling of diffuse gas infall (Sec.A.1.2).

A.2 Discussion

In order to match the resolution of SH and HAGN in galaxies, the EH simulation is limited to kpc-scale resolution, so the real compactness of UC galaxies could be under-estimated as they are as compact as the resolution limit allows. Zoom-in simulations will be required to make robust assessment of their size distribution. Nevertheless, the population of UC galaxies in EH is tightly associated with specific formation histories dominated by major mergers of low-mass progenitors, compared to larger galaxies in the simulation.

To further probe the effect of feedback in compact galaxy formation, we used the Horizon-AGN suite of simulations from Chabanier et al. (2020). These simulations are run with extreme feedback parameters leading to barely realistic variations of the black hole-to-stellar mass ratio, yet the average galaxy size at fixed stellar mass changes by less than 10%, confirming that feedback is not a key driver of the formation of UC galaxies in EH. This also suggests an optimistic note on computational galaxy formation : the increase in resolution, without adjusting or modifying subgrid recipes, appears more important than sub-grid physics and can succeed in explaining key properties such as the mass-size relation.

We have analyzed so far the compactness of galaxies independently from their star formation activity. As expected for galaxies in the 10^{10}-10^{11} $\mathrm{M}_\odot$ stellar mass range at $z = 2$, both NUC and UC galaxies are mainly star-forming galaxies on the MS. There is nevertheless a clear trend for compact galaxies to have relatively low specific star formation rates (sSFR, Fig. A.5). The majority of UC galaxies lie on the low-sSFR end of the MS, as observed for blue nuggets (Barro

---

[3] similar criteria hold for the fourth and fifth most massive progenitors and are also valid when the same analysis is performed at $z = 4$.

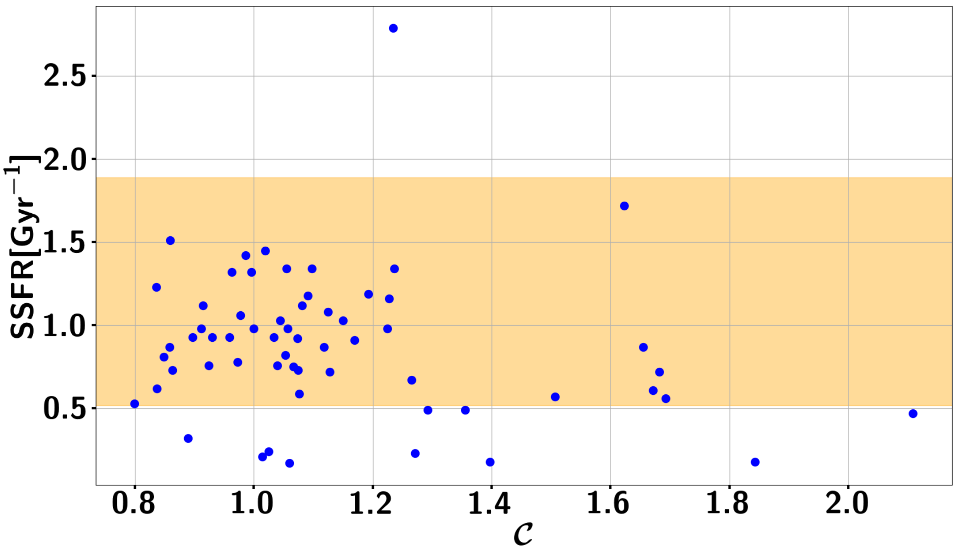

FIGURE A.5 – *Specific Star Formation Rate (SSFR) as a function of compactness ($\mathcal{C}$ for EH galaxies at $z = 2$. The shaded area defines the Main Sequence following Schreiber et al. (2017).*

et al., 2017). The relatively low sSFRs of UCs, as well as a tentative excess of galaxies below the MS among UCs compared to NUCs, are consistent with the idea that these objects are undergoing quenching through gas exhaustion and/or feedback (Tacchella et al., 2016).

The number of UC galaxies in EH (10 objects in $(50\,\mathrm{Mpc/h})^3$) is consistent with the number density of compact SFGs (see Introduction). The EH volume is too small to firmly probe the formation of massive compact ETGs at z=2, as statistically about one such object is expected in this volume, but the excess of low-mass progenitors in EH is already present at z=4 (Fig. A.1) and could explain the early formation of such compact ETGs. There is indeed one galaxy in EH with M_*=1.2× $10^{11}\,\mathrm{M}_\odot$ and compactness $\mathcal{C}$=1.29 (almost UC in our definition), with a low SSFR=0.23 Gyr^{-1} (a factor 7 below the MS), a low gas fraction of 11% (within $3R_e$), and a Sersic index of 3.6 at z=2. This galaxy continues to quench into a compact ETG by redshift $z \simeq 1.8$, with SSFR=0.13 Gyr^{-1}, M_*=1.7× $10^{11}\,\mathrm{M}_\odot$, and R_e=4.0 kpc at $z \simeq 1.8$, thus lying close to the mass-size relation of ETGs at $z = 1.75$ from van der Wel et al. (2014). This candidate compact ETG does also form through major mergers of low-mass progenitors : its two main progenitors at z=4 contain 11 and 8% of its stellar mass, respectively.

To compare the environment of UC and NUC galaxies, we study the large-scale structure of the EH simulation with the persistent skeleton approach (Sousbie, 2011) using the `DISPERSE` code (Sousbie, 2013). The full skeleton is shown in Fig. A.6. Topological persistence can be used to characterise the significance of the structures depending on the local level of noise. Persistence levels from 3 to 8 σ are used to investigate different scales and prominences of the corresponding cosmic web (the larger the persistence threshold the lower the total number of filaments but those filaments are the most prominent ones). The 3σ threshold insures that we are not dominated by noise. On the other hand, above 8σ very few structures are left. Fig. A.6 shows an intermediate situation with a persistence threshold at 7σ. In the remainder of this section, we will focus on two cases : i) a low-persistence skeleton (3σ) where by definition all massive galaxies reside inside filaments and nodes which we can characterise (in terms for instance of density, connectivity, etc) ; and ii) a high-persistence skeleton (8σ) which only selects the dominant filaments in the simulation box and for which we can measure the distance from galaxies to their closest filament or node.

At high persistence, the skeleton is sparse, dominated by a few dense and extended filaments. UC and NUC galaxies both lie close to such filaments, as expected for massive galaxies in general, but the galaxies that lie closest to these dense filaments and their nodes are never UC (Fig. A.7, top panel). Instead, UC galaxies tend to lie in intermediate-density filaments, as shown by the analysis of the closest filaments in a lower-persistence skeleton analysis (Fig. A.7, bottom panel). This is consistent with the previous results on the merger history of UC galaxies, as objects in the densest regions of the main filaments are expected to form their main progenitor early-on and subsequently grow by minor mergers and/or diffuse accretion. UC galaxies nevertheless still do form in dense regions and none is found in low-density filaments where smooth accretion would dominate over mergers (Fig. A.7 right panel and Fig. A.6 for a visualisation).

Hence, UC galaxies are expected to be found in relatively dense environments, but not in the very densest filaments and nodes. Galaxies in the densest regions of the cosmic web are expected to be rarely ultra-compact at $z \sim 2$, yet could undergo ultra-compact phases at higher redshift if their early formation involves major mergers of numerous low-mass progenitors.

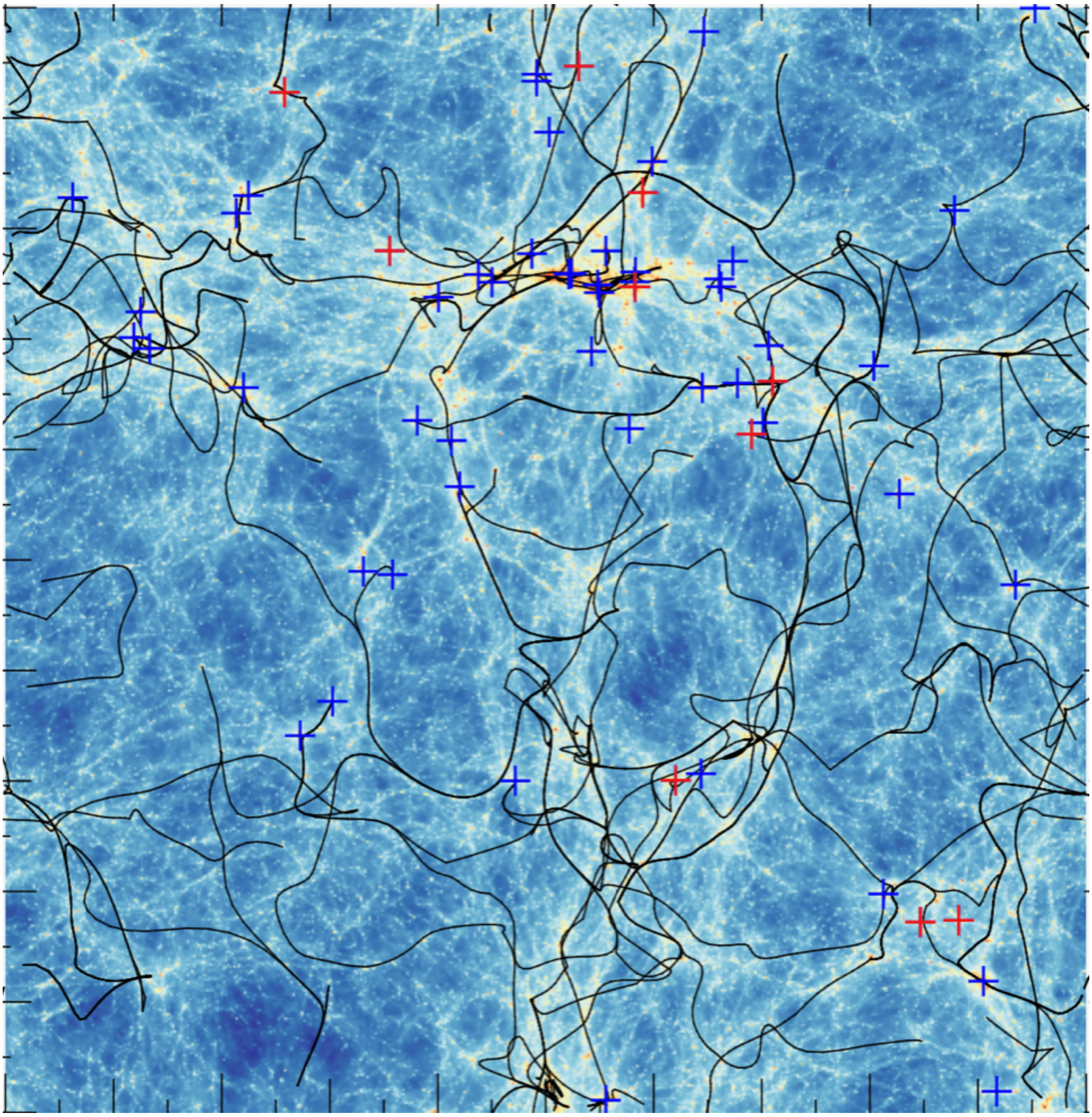

Figure A.6 – *Projected EH skeleton of the 50 Mpc h^{-1} box with a 7σ persistence level at $z = 2$. Crosses indicate the projected position of massive galaxies, with red crosses for UC and blue crosses for NUC.*

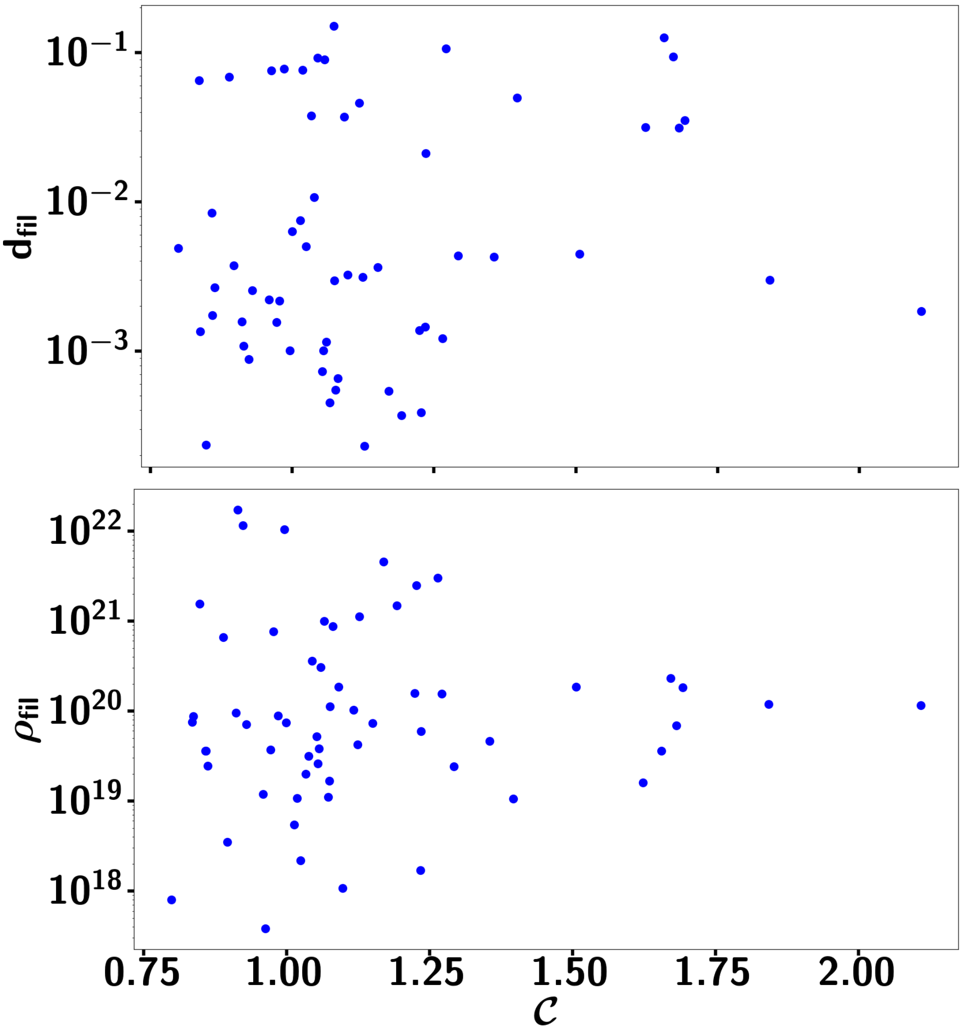

FIGURE A.7 – ***Top :*** *Distance to the closest filament* d_{fil} *(in box size units) of the* 8σ *sparse skeleton for all EH massive galaxies as a function of their compactness.* ***Bottom :*** *Density in the closest filament* ρ_{fil} *(obtained by DTFE from a mass-weighted Delaunay tessellation of the galaxy catalogue) of the* 3σ *dense skeleton for all EH massive galaxies as a function of their compactness. For UC galaxies, exclusion zones are clearly visible at small distance to the filaments and in the very low and very high density regions, compared to the NUC.*

B

Résumé Substantiel

B.1 Introduction

La théorie englobant l'hypothèse d'un univers statistiquement homogène et isotrope aux grandes échelles et le modèle du Big Bang chaud dans lequel les petites fluctuations primordiales de densité ont grandi par le mécanisme d'instabilité gravitationnelle pour donner naissance aux grandes structures observées aujourd'hui constitue le modèle ΛCDM qui permet d'expliqur un grand nombre d'observations indépendantes sur une large gamme d'échelles et de temps cosmiques. Le nom provient du contenu énergétique mesuré, avec 5% de matière ordinaire, ou matière baryonique, qui compose le gaz, les planètes ou les étoiles, 25% de matière noire composée de particules froides n'interagissant pas avec la lumière et seulement très peu avec la matière bayronique, et enfin 70% d'énergie noire, responsable de l'expansion accélérée de l'univers qui peut-être interprétée comme une constante cosmologique, Λ, dans les équations de la relativité générale d'Einstein. Cependant, ce modèle n'est pas parfait et la cosmologie moderne doit encore faire face à de nombreux challenges.

Tout d'abord, ce modèle est un modèle phénomènologique qui contraint les propriétés de la matière noire et de l'énergie noire mais ne donne pas d'indication sur leur nature. De plus, nous sommes entrés dans l'ère de cosmologie de précision signifiant que la précision des mesures des paramètres cosmologiques ont atteint le pourcent. Bien que la plupart des mesures convergent vers le modèle ΛCDM, certaines sondes indépendantes sont en tension à plus de 5σ. Ces différences pourraient provenir de biais non identifés dans les analyses de données ou pourraient indiquer la nécessité d'une physique au-delà du modèle standard. Dans le cadre de ma thèse, je m'intéresse à deux problèmes fondamentaux du paradigme standard : les neutrinos massifs et les tensions aux petites échelles du modèle de matière noire froide.

B.1.1 Les neutrinos massifs

Les neutrinos sont les particules élémentaires du fluide cosmique les plus abondantes de l'univers après les photons et interagissent très peu avec la matière bayronique. Par analogie avec les leptons électroniquement chargés (l'électron, le muon et le tauon), ils existent sous trois saveurs : le neutrino électronique (ν_e), le neutrino muonique (ν_μ) et le neutrino tauique (ν_τ). Les mesures d'oscillations des neutrinos, mécanisme découvert en 2000 (Fukuda et al., 1998; Kajita, 1999; Ahmad et al., 2001), ont prouvé qu'au moins deux des trois neutrinos sont massifs contrairement à la prédiction du modèle standard de la physique des particules qui les décrit avec des masses nulles, ce qui indique la nécessité d'une physique au-delà du modèle standard. Du point de vue de la cosmologie ceci pose un problème majeur car nos avons atteint un niveau de précision suffisant pour être sensibles à l'impact de la masse des neutrinos qui peut significativement modifier l'histoire de la formation des structures par rapport à un modèle d'univers avec des neutrinos non massifs.

Les mesures d'oscillations des neutrinos ne permettent pas de sonder les masses individuelles mais plutôt les deux différences des carrés des masses, $\delta m^2_{21} = (7.53 \pm 0.18) \times 10^{-5}\mathrm{eV}^2 > 0$ et $|\Delta m^2_{32}| \simeq |\Delta m^2_{31}| = (2.44 \pm 0.06) \times 10^{-3}\mathrm{eV}^2$. Néanmoins, une incertitude demeure sur le signe de Δm^2_{32}, i.e. nous ne savons pas lequel des trois neutrinos est le plus léger et la valeur de sa masse. Ceci donne lieu à deux scénarios de masse, résumés en figure B.1 :

- le scénario de masse normale si $m_3^2 > m_2^2$ avec une somme des masses minimale $\sum m_\nu = 0.06$ eV,
- le scénario de masse normale si $m_2^2 > m_3^2$ avec une somme des masses minimale $\sum m_\nu = 0.1$ eV.

Nous en déduisons que la somme des masses des neutrinos est un paramètre clé pour discriminer ces hierarchies de masse.

$\sum m_\nu$ peut-être sondé en utilisant la cosmologie, et plus précisément la formation des structures via le spectre de puissance de la matière qui mesure l'intensité des fluctuations de densité de la matière en fonction de l'échelle. Nous définissons le champs de densité de contraste de la matière comme

$$\delta_m(\mathbf{x}) = \frac{\rho_m(\mathbf{x}) - \langle \rho_m \rangle}{\langle \rho_m \rangle}, \tag{B.1}$$

avec ρ_m la densité de matière. Le spectre de puissance est alors exprimé comme

$$P_m(k) = \left\langle |\tilde{\delta}_m(k)|^2 \right\rangle, \tag{B.2}$$

où $\tilde{\delta}_m$ est la transformé de Fourier de δ_m.

La physique des particules donne déjà de très bonnes contraintes avec $0.06 \text{ eV} \leq \sum m_\nu \leq 3$ eV, ce qui nous permet de savoir que les neutrinos sont assez légers pour avoir connu une période dite relativiste pendant laquelle ils se comportaient comme des photons. Et lorsque la température de l'univers a suffisamment diminué, leur vitesse devient négligeable par rapport à leur masse et ils se comportent comme de la matière noire mais avec une dispersion de vitesse non négligeable, contrairement à la matière noire froide. Grâce à leur pression de radiation, les neutrinos peuvent se propager librement sur de grandes distances. Plus leur masse est petite, plus ils peuvent parcourir des distances élevées et empêchent des structures plus larges de s'effonder

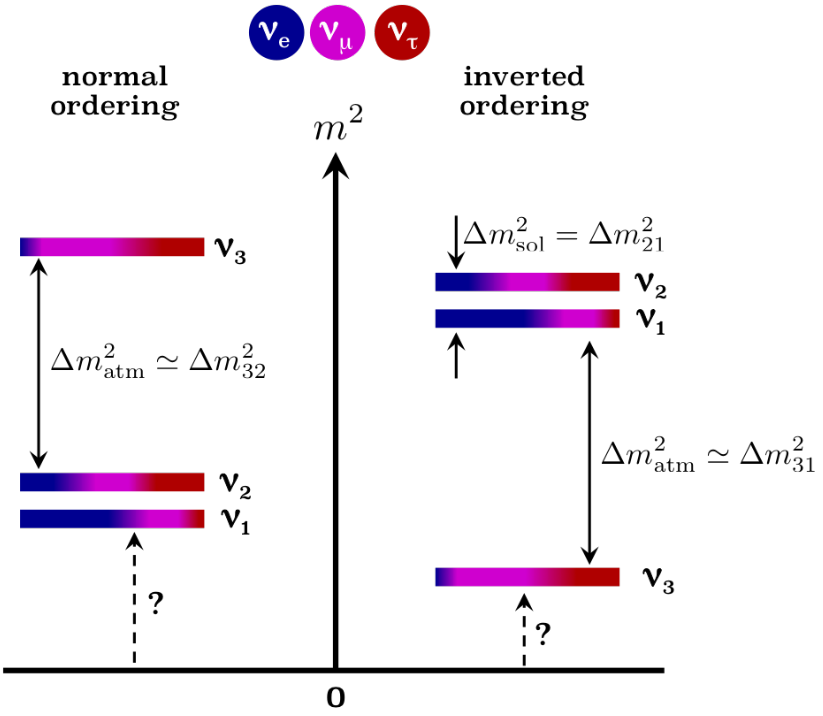

Figure B.1 – *Hiérarchie de masse normale (gauche) and inverse (droite). Credits : Julien Baur*

par instabilité gravitationnelle conduisant à un déficit de puissance aux petites échelles, c'est-à-dire aux grands nombres d'onde $k \gg k_{nr}$, où k_{nr} correspond au nombre d'onde de la distance de libre propagation d'un neutrino qui est exprimé comme

$$k_{\rm nr} \approx 0.018\ \Omega_m^{1/2} \left(\frac{m}{1\ \rm eV}\right)^{1/2}\ \rm h\ Mpc^{-1}. \tag{B.3}$$

La théorie linéaire prédit que pour $k \gg k_{nr}$ on a

$$\frac{P^{f_\nu \neq 0}}{P^{f_\nu = 0}} \approx -8 f_\nu, \tag{B.4}$$

où $f_\nu = \Omega_\nu / \Omega_m$ est le paramètre d'abondance des neutrinos, $P^{f_\nu \neq 0}$ le spectre de puissance de la matière avec des neutrinos massifs et $P^{f_\nu = 0}$ le spectre de puissance de la matière avec des neutrinos non massifs. La prédiction de l'impact des neutrinos massifs sur le spectre de puissance de la matière à des échelles si petites doit être calculée à l'aide de simulations numériques prenant en compte précisément les non linéarités du champ de densité, mais la théorie linéaire prouve tout de même que $\sum m_\nu$ peut être efficacement contraint avec la cosmologie.

B.1.2 Le modèle de matière noire tiède

Le modèle de matière noire froide présente des tensions avec les observations aux petites échelles. Par exemple, il prédit l'existence de nombreuse galaxies naines qui n'ont pas été observées en nombre suffisant. Il prédit également que la distribution de densité des halos de matière noire forme des pics beaucoup plus marqués que ce qui est observé. Tous ces problèmes ajoutés à l'absence de détéction directe de particules de matière noire froide pourraient plutôt indiquer un problème dans notre hypothèe de matière noire dite froide.

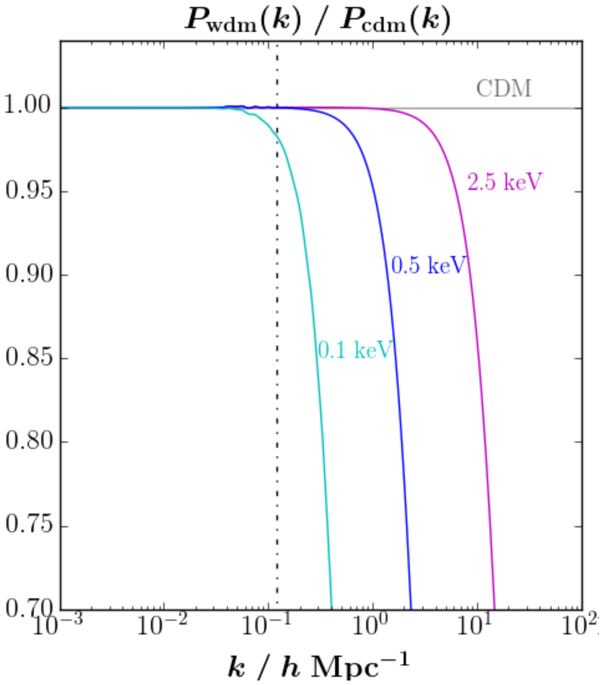

Figure B.2 – *Approximation analytique de la fonction de transfert définie comme* $T^2(k, z = 0) = P_{wdm}(k, z = 0)/P_{cdm}(k, z = 0)$*, où* P_{wdm} *est le spectre de puissance pour un modèle de matière noire tiède et* P_{cdm} *le spectre de puissance pour un modèle de matière noire froide. Les différentes couleurs encodent différentes masses de particules de matière noire : 0.1 keV (cyan), 0.5 keV (bleu) et 2.5 keV (magenta). La région à droite de la ligne noire pointillée indique les échelles sondées par la forêt Ly*α*, la sonde cosmologique utilisée dans cette thèse. De Baur et al. (2016).*

De nombreux modèles théoriques visent à concilier les prédictions du modèle aux grandes et petites échelles avec les observations, comme par exemple le modèle de matière noire tiède dans lequel les particules de matière noire possède une dispersion de vitesse non négligeable, ce qui leur confère une distance de libre propagation comparable à taille d'une galaxie naine, $\lambda_{FS} \sim 0.1\ \mathrm{h}^{-1}\cdot\mathrm{Mpc}$, sous laquelle elles suppriment la formation des structures. Par analogie avec la masse des neutrinos cosmologiques, le spectre de puissance de la matière permet d'étudier la probabilité d'un modèle de matière noire tiède en sondant une suppression de puissance aux petites échelles, comme le montrent les prédictions linéaires de la figure B.2.

J'ai considéré dans mon travail de thèse deux candidats comme particules de matière noire tiède qui ont le même impact sur la formation des structures. Les premiers sont les reliques thermiques qui sont des particules hypothètiques très massives qui auraient été initialement à l'équilibre thermique avec le reste du plasma cosmique. Elles se seraient découplées de ce plasma lorsque l'univers était très jeune, bien avant le découplage des neutrinos, donc leur température est réduite d'un facteur $T_x/T_\nu = \alpha < 1$ étant donné qu'ils ne recoivent pas l'énergie libérée des différentes annihilations de cette époque. Etant donné qu'elles étaient couplées aux photons avant le découplage des neutrinos, elle participe comme contribution additionnelle au nombre effectif de neutrinos, $N_{\rm eff}$, tel que

$$N_{\rm eff} = 3.046 \pm \Delta N_{\rm eff} \tag{B.5}$$

avec

$$\Delta N_{\rm eff} = \left(\frac{T_x}{T_\nu}\right)^4 . \tag{B.6}$$

Le nombre d'onde au dessus duquel le spectre de puissance commence à dévier du spectre de puissance d'un modèle de matière noire froide, k_0, peut être écrit

$$k_0 = 0.24 \left(\frac{m_x}{1 \text{ keV}} \right)^{-0.83} (\Delta N_{\text{eff}})^{0.83/4} \left(\frac{\omega_x}{0.25(0.7)^2} \right)^{-0.16} \text{ Mpc}, \tag{B.7}$$

où $w_x = \Omega_x / \Omega_m$ est l'abondance des reliques thermiques.

Les deuxièmes candidats étudiés sont les neutrinos stériles, des neutrinos hypothètiques plus massifs que les neutrinos standards et qui n'interagissent que par gravité, à l'opposé des neutrinos standard qui sont également sensibles à la force d'interaction faible, ce qui fait d'eux d'excellent candidats à la matière noire mais les rend également très difficile à détecter. Plusieurs mécanismes de production existent, avec différents niveau de complexité. Nous avons considéré ici un modèle générique simple de production de neutrinos stériles dits non-résonants, le mécanisme de Dodelson-Widrow (Dodelson and Widrow, 1994), dans lequel ils émergent via des oscillations avec les neutrinos standards lorsque l'univers était très jeune, tel que $T < 1$ GeV.

Etant donné les similarités entre les distributions de l'espace des phases des reliques thermiques et des neutrinos stériles non résonants, il existe une correspondance simple entre la masse d'un relique thermique, m_x, et la masse d'un neutrino stérile non résonant, m_{ν_s}, qui produit le même impact sur le spectre de puissance de la matière,

$$\omega_{dm} \cdot m_{\nu_s}^3 \propto m_x^4. \tag{B.8}$$

B.1.3 La forêt Lyman-alpha

Afin de déterminer la distribution de matière totale dans l'univers, δ_m, nous devons avoir recours à une sonde visible qui trace plus ou moins fidélement la distribution de matière. Dans le cadre de ma thèse, j'utilise la forêt Lyman-α (Lyα) comme observable cosmologique car elle sonde des échelles particulièrement petites, de l'ordre du Mpc, et des époques de l'univers relativement jeune, là où l'impact des particules relativistes sur la formation des structures est très marqué.

La forêt Lyα constitue la collection de raies d'absorptions Lyα visibles dans les spectres d'objets lointain, tel que les quasars. Les émissions de Lyman correspondent aux raies d'émission dans l'UV de l'hydrogène neutre provoquée par la transition électronique depuis un état excité $n \geq 2$ à l'état principal $n = 1$. La première transition, depuis $n = 2$ à $n = 1$, correspond à la raie Lyα, qui émet (ou absorbe) à la longueur d'onde $\lambda_{Ly\alpha} = 1,215.17$ Å.

Les photons émis par des quasars se propagent dans un univers en expansion et leur longueur d'onde, λ, est diluée par rapport à la longueur d'onde dans son référentiel lors de l'émission, λ_0, tel que $\lambda = (1 + z)\lambda_0$. Le spectre d'émission du quasar se retrouve ainsi constamment décalé vers le rouge au fur et à mesure que la lumière se propage. L'absorption Lyα a lieu en continu par des nuages d'hydrogène neutre du milieu intergalactique à la longueur d'onde $\lambda_{Ly\alpha}$ dans le référentiel de l'absorbeur ce qui donne naissance à la série de raies d'absorption Lyα à gauche de la raie d'émission Lyα, appellé forêt Lyα. Le décalage spectral entre l'absorption et la raie d'émission renseigne sur la localisation spatiale du nuage d'hydrogène neutre le long de la ligne de visée, et la profondeur de la raie d'absorption renseigne sur l'amplitude de la surdensité. La figure B.3 illustre cette formation des forêts d'absorptions.

les simulations numériques montrent que la forêt Lyα provient des fluctuations de densité et de vitesse dans le milieu intergalactique. A hauts redshifts, $z \geq 2$, l'effondrement gravitationnel d'une grande majorité du gaz du milieu intergalactique n'a atteint que le régime moyennement non linéaire tel que $\delta \leq 10$. La dynamique du milieu intergalactique est au premier ordre assez

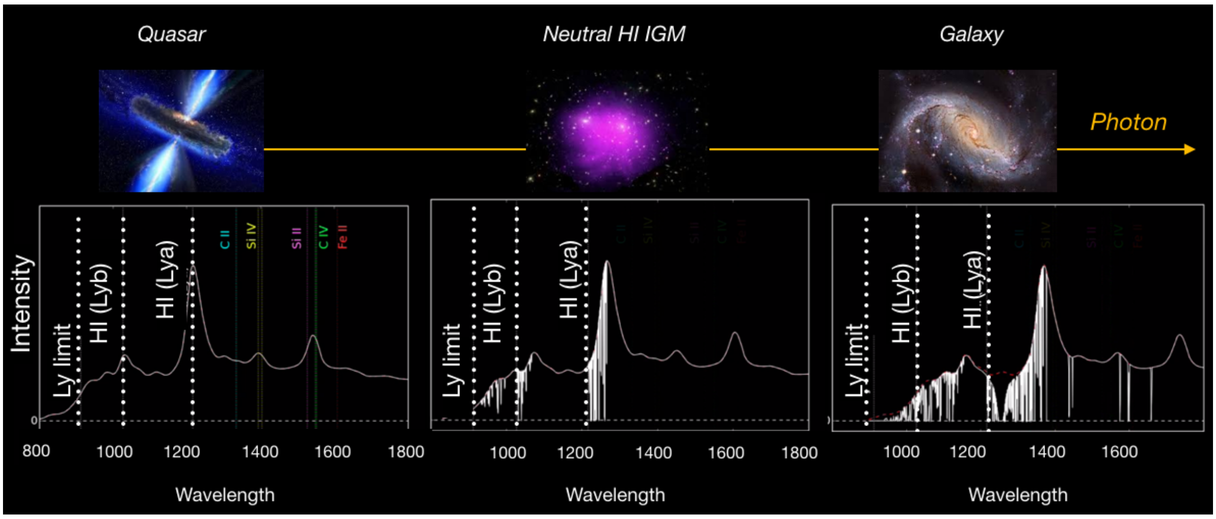

FIGURE B.3 – *Illustration de l'impact d'absorbeurs situés sur la ligne de visée d'un qusar sur le spectre observé. Je montre le spectre du quasar à différentes localisations le long de la ligne de visée : à l'émission (gauche), dans le milieu intergalactique (milieu) et proche d'une galaxie (droite). Adapté de la vidéo d'Andrew Pontzen.*

simple, et émerge de la combinaison de l'effondrement gravitationnel des grandes structures, dominées par la matière noire, et de la pression baryonique aux petites échelles. Ainsi, la forêt Lyα est un traceur de la matière noire, et est donc sensible à la cosmologie.

B.2 Mesure du spectre de puissance de la forêt Lyα

Dans le cadre de ma thèse, j'ai participé à l'un des plus ambitieux programmes d'observations, le programme eBOSS (extended-Baryon Oscillation Spectroscopic Survey), qui utilise le téléscope du SDSS (Sloan Digital Sky Survey) de 2.5m situé à l'observatoire de Apache Point au Nouveau Mexique, aux Etats-Unis. Ce téléscope est en opération depuis 20 ans maintenant et a observé plus de deux millions de galaxies et quasars, dont environ 200,000 quasars à haut redshift dédiés aux études des forêts Lyα. Je présente dans cette partie l'analyse de données des spectres de puissance des forêts Lyα d'un échantillon de 43,751 quasars issu de la 14$^{\text{ème}}$ *data release* (DR14) publié dans Chabanier et al. (2019).

B.2.1 Méthode

Le champ de fluctuations d'absorption du flux est défini comme

$$\delta(\lambda) = \delta_{\text{Ly}\alpha}(\lambda) + \delta_{\text{background}}(\lambda) + \delta_{\text{noise}}(\lambda), \tag{B.9}$$

où $\delta_{\text{Ly}\alpha}(\lambda)$ rend compte des absorptions Lyα, $\delta_{\text{background}}(\lambda)$ des autres absorptions, qui peuvent avoir une faible séparation en longueur d'onde et sont donc dites corrélées aux absorptions Lyα, comme Si III, ou une large séparation en longueur d'onde et sont donc dites décorrélées, comme C IV, et finalement $\delta_{\text{noise}}(\lambda)$, le bruit poissonien, non corrélé aux absorptions physiques. Le spectre de puissance du flux est défini comme le carré de la transformé de Fourier du champ de fluctuations d'absorptions, $|\mathcal{F}(\delta)|^2$. Notre estimateur du spectre de puissance du flux est donc donné par

$$P_{\text{raw}}(k) = (P_{\text{1D}}(k) + P_{\text{metals}}) \cdot W^2(k, R, \Delta v) + P_{\text{noise}}(k), \tag{B.10}$$

où P_{1D} est l'estimateur du spectre de puissance des absorptions Lyα et des absorptions corrélées aux absorptions Lyα, P_{metals} le spectre de puissance des absorptions décorrélées des absorptions Lyα, P_{noise} le spectre de puissance du bruit poissonien et W est la fonction fenêtre correspondant à la reponse spectrale du spectrographe. Cette-ci dépend de la largeur d'un pixel et de la résolution du spectrographe R tel que

$$W(k, R, \Delta v) = \exp\left(-\frac{1}{2}(kR)^2\right) \times \frac{\sin(k\Delta v/2)}{(k\Delta v/2)}\,. \tag{B.11}$$

B.2.2 Incertitudes systématiques

Par rapport à l'analyse précédente, utilisant les données DR9 et publiée dans Palanque-Delabrouille et al. (2013), nous avons trois fois plus de quasars. Ceci a pour résultat de significativement réduire les incertitudes statistiques mais nous rend également bien plus sensibles aux incertitudes systématiques. Ainsi, pour ne pas biaiser nos interpreations cosmologiques, j'ai d'abord amélioré la quantification des erreurs systématiques mais j'ai également étudié des nouvelles sources d'erreur. Nous avons dans ce travail identifié au total huit incertitudes systématiques liées à :

- Mesure du spectre continu du quasar
- Mesure du spectre de puissance du bruit
- Mesure de la résolution du spectrographe
- Mesure du spectre de puissance des absorptions décorrélées des absorptions Lyα
- L'effet du masquage des lignes de ciel
- L'effet du masquage des DLA
- L'effet d'incomplétude du catalogue de DLAs (Damped Lyα systems)
- L'effet d'incomplétude du catalogue de BALs (Broad Absorption Lines)

Les figures B.4 et B.5 montrent les rapports des huit incertitudes sytématiques identifiées par rapport aux incertitudes statistiques ainsi que leur valeur absolue. Celle-ci sont décorrélées les unes des autres et peuvent donc être ajoutées en quadrature.

Notre méthode de mesure de spectre continu du quasar introduit des corrélations résultant à un excés de puissance aux grandes échelles que nous corrigeons en utilisant des données synthétiques, des *mocks*, et nous ajoutons une incertitude de 30% sur cette correction au budget d'erreur systématique.

La précision avec laquelle est mesurée le bruit dans le *pipeline* automatique de SDSS n'est pas suffisante étant donné la précision nécessaire pour contraindre l'impact des particules relativistes sur la formation des structures avec le $P_{\text{Ly}\alpha}$. En effet, le bruit poissonien impacte principalement la mesure de l'amplitude du spectre de puissance qui est hautement dégénérée avec l'amplitude de la supression causée par la libre propagation des particules relativistes. Il est donc fondamental d'améliorer la mesure du bruit par rapport à celle fournie par le *pipeline*. Nous construisons un spectre denué de tout signal astrophysique en utilisant les expositions individuelles, dont le carré de la transformé de Fourier est notre estimateur du spectre de puissance du bruit qui apparaît être indépendant de l'échelle, comme attendu pour le spectre de puissance d'un bruit blanc. Cependant, P_{noise} est supposé être l'asymptote de P_{raw} aux petites échelles mais ce n'est pas le cas aux petits redshifts. On trouve une différence au pire de 5% pour les trois plus petits bins en redshifts et aucune différence pour les autres. On introduit donc une incertitude sur la mesure de P_{noise} de 30% de la plus grande différence entre l'asymptote de P_{raw} et P_{noise} pour tous les redshifts.

L'une des systématiques dominantes aux petites échelles provient de la connaissance de la résolution spectrale du spectrographe. Celle ci a été mesurée dans SDSS en utilisant des lampes à arc. Plusieurs études (Smee et al., 2013; Palanque-Delabrouille et al., 2013) ont montré que la précision de $\Delta R/R$ sur la mesure de R est d'environ 2%. L'incertitude systématique sur $P_{1\mathrm{D}}(k)$ liée à la résolution est donc donnée par $P_{1\mathrm{D}}(k) \cdot (2k^2 R\Delta R)$. La dépendance en k^2 rend cette incertitude plus importante aux petites échelles (grand k). Aussi, il est important de noter que la resolution moyenne, $< R >$, varie avec le redshift, de 81 s/km à 65 s/km, avec la plus haute valeur pour les plus petits redshifts, rendant l'incertitude plus élevée à bas redshifts.

Les spectres de puissance des absorptions décorrélées des absorptions Lyα, P_{metals}, sont mesurés dans l'intervalle de longueur d'onde $1,270\text{Å} < \lambda_{\mathrm{RF}} < 1,380\text{Å}$, qui contient les contributions de tous les métaux avec longeur d'onde d'absorptions $\lambda_{\mathrm{RF}} > 1,380\text{Å}$ tel que Si IV ou C IV. Nous propageons son incertitude statistique comme incertitude systématique.

La qualité des pixels affectés par des émissions de ligne de ciel est drastiquement réduite à cause d'une forte augmentation du bruit, ce qui nous conduit à masquer ces pixels dans l'analyse de données. L'effet du masquage introduit un biais qui dépend à la fois du redshift et de l'échelle qu'on estime avec les *mocks*. Comme pour l'incertitude sur la mesure du spectre continu du quasar, nous ajoutons une incertitude de 30% sur cette correction au budget d'erreur systématique.

Les absorptions saturées de DLAs ne sont pas causées par des nuages d'hydrogène neutre dans le milieu intergalactique mais plutôt par des reservoirs de gaz situés en périphérie de galaxies qui biaisent nos mesures en produisant des corrélations additionnelles si nous ne les écartons pas de notre jeu de données, en particulier aux grandes échelles, car ceux ci ne sont pas simulés dans nos prédictions théoriques. De la même manière que pour les pixels affectés par les lignes de ciel, nous masquons les pixels affectés par les absorptions de DLAs, tandis que l'analyse BOSS-Lyα (Palanque-Delabrouille et al., 2013) rejetait complétement les spectres affectés par de tels systémes et réduisait ainsi significativement la taille de l'échantillon analysé. En effet, environ 5% de 43,751 spectres contienne des DLAs. Pour identifier les DLAs, nous utilisons l'algorithme de detection automatique de Noterdaeme et al. (2012) qui fitte des profils de Voigt sur les absorptions des forêt Lyα. J'appelerai ce catalogue N12. L'effet de masquage des pixels affectés par les DLAs est également étudié avec les *mocks* qui montrent un biais inférieur à 1%, peu dépendant de l'échelle et qui augmente avec le redshift, ce qui est cohérent avec l'augmentation de la fraction des spectres contenant des DLAs avec le redshift. Nous ajoutons une incertitude de 30% sur cette correction au budget d'erreur systématique.

Pour estimer l'impact de DLAs résiduels non masqués dans notre échantillon, causé par une éventuelle incomplétude de l'algorithme de Noterdaeme et al. (2012), nous utilisons un catalogue alternatif, P18, publié dans Parks et al. (2018) qui utilise un réseau de neurones. P18 contient 4,419 DLAs sur notre échantillon de spectres quand N12 en contient 2,105. La comparaison entre les spectres de puissance où nous masquons tous les DLAs contenus dans l'union de N12 et P18 aux spectres de puissance où nous ne masquons que le DLAs de N12 montre une augmentation de puissance aux grandes échelles. Nous ajoutons une incertitude de 30% sur ce rapport au budget d'erreur systématique pour rendre compte de l'incomplétude de notre catalogue de DLAs.

De manière similaire, les BALs sont produits par des absorbeurs situés en périphérie du disque d'accrétion du quasar qui baisent fortement les mesures car nous ne les simulons pas dans nos prédictions théoriques. Nous rejetons tous spectres affectés par de tels absorbeurs identifiés avec

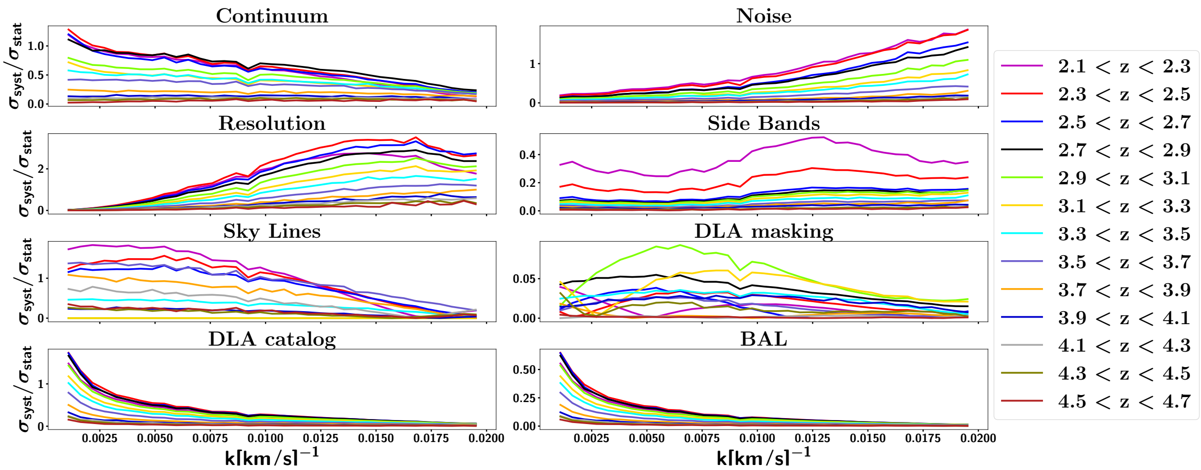

FIGURE B.4 – *Rapports des incertitudes systématiques sur les incertitudes statistiques en fonction du redshift et du nombre d'onde. Du haut vers le bas et de la gauche vers la droite sont illustrés les rapports des incertitudes causées par l'estimation du continu du quasar, de la mesure du niveau de bruit, de la connaissance de la résolution spectrale, de la mesure du spectre de puissance des absorptions décorrélées aux absorptions Lyα, de l'effet du masquage des lignes de ciel, de l'effet du masquage des DLAs, de l'incomplétude des catalogues de DLAs et de BALs.*

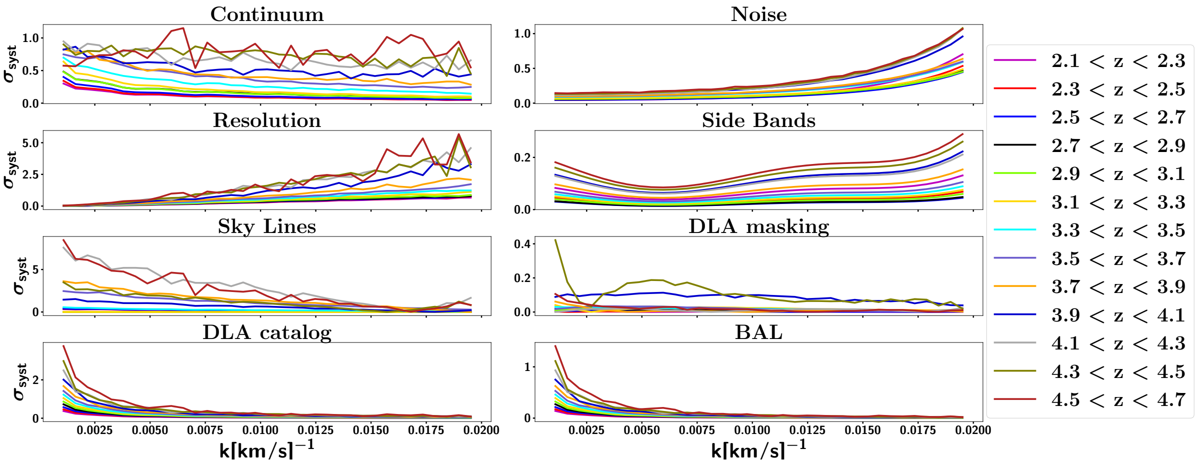

FIGURE B.5 – *Valeurs absolues des incertitudes systématiques en fonction du redshift et du nombre d'onde. Le format est la même que pour la figure B.4.*

des *Balnicity Index* (BI) supérieurs à 0 dans le catalogue de quasars DR14Q. Nous considérons que la procédure automatique identifie presque parfaitement 80% des BALs, soit ceux avec BI > 170 km/s, mais qu'elle peut en manquer pour les 20% moins visibles restants. Nous ajoutons comme incertitude systématique liée à l'effet de BALs résiduels dans notre échantillon le rapport $P_{BI>0}(k)/P_{BI>170}(k)$.

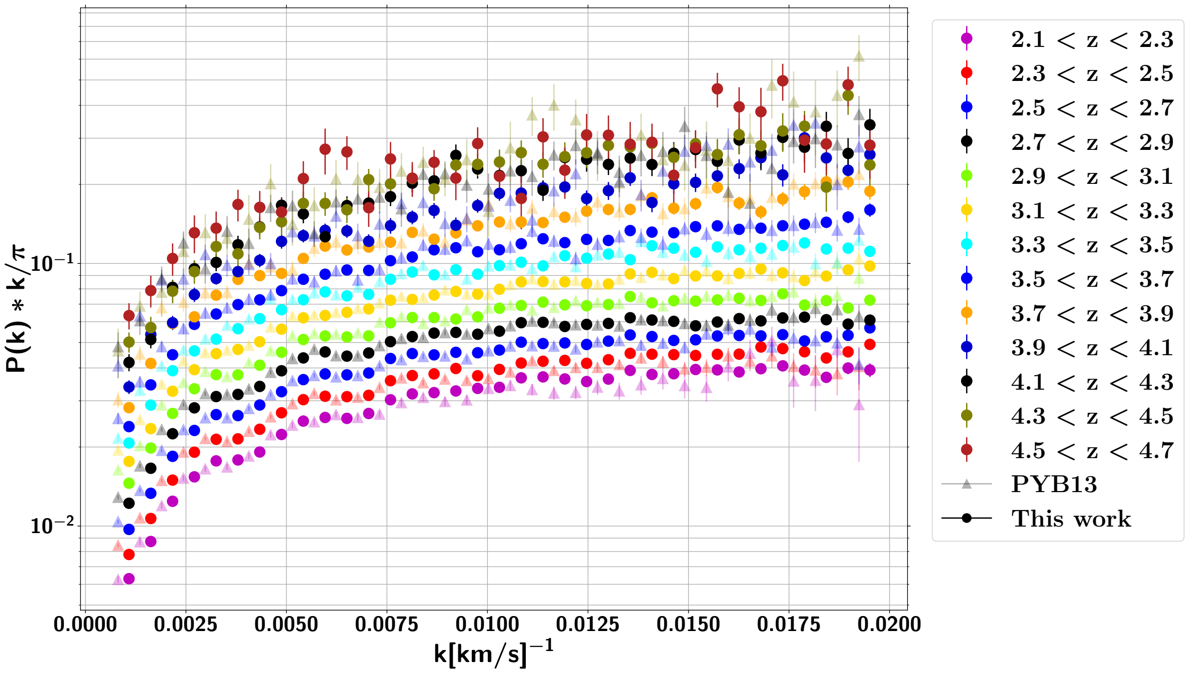

FIGURE B.6 – *Les spectres de puissance en 1D des forêts Lyα. Les mesures de l'analyse de données de Palanque-Delabrouille et al. (2013) (PYB13) sont montrées en couleurs plus clair et légérement décalées vers la gauche pour une meilleure lisibilité. Les barres d'erreurs prennent seulement en compte les barres d'erreur statistiques.*

B.2.3 Résultats

En utilisant la procédure décrite dans la section B.2.1, nous mesurons le spectre de puissance des forêts Lyα de l'échantillon de 43,751 spectres de quasars de DR14 en 13 bins en redshifts avec $2.2 \leq z_{Ly\alpha} \leq 4.6$ avec chacun 35 nombres d'onde tel que 10^{-3} s/km $\leq k \leq 0.02$ s/km. Les spectres de puissance sont montrés en figure B.6 en comparaison de l'analyse de données sur DR9 publiée dans Palanque-Delabrouille et al. (2013). Les deux analyses sont en très bon accord et ne montrent aucune déviation significative. On observe une forte réduction des erreurs, d'un facteur 3 voire 4, en particulier aux petits redshifts où nous atteignons une précision inférieure au pourcent. Les incertitudes statistiques sont réduites d'un facteur 4 grâce à l'augmentation d'un facteur 3 sur le nombre de quasars et d'une séléction plus stricte de l'échantillon tandis que les incertitudes systématiques ont augmenté d'un facteur 2 en raison d'une analyse plus fine et rigoureuses des erreurs et de leur sources. Finalement, nous avons un bin en redshift supplémentaire, $z = 4.6$, qui est particulièrement utile pour l'étude des modèles de matière noire tiède étant donné que ce redshift est moins affecté par la croissance non linéaire des structures.

B.3 Prédictions théoriques issues de simulations hydrodynamiques

La forêt Lyα sonde des échelles relativement non linéaires et met en jeux une physique très riche, qui émerge de la combinaison complexe entre l'évolution des grandes structures et de la physique baryonique des petites échelles et nous devons donc avoir recours à des simulations hydrodynamiques cosmologiques pour construire nos prédictions théoriques. Les mesures du spectre de puissance des forêts Lyα atteignent un niveau de précision sans précédent, inférieur au pourcent aux petits redshifts. Néanmoins, nous ne pouvons faire des intérprétations cosmologiques robustes que si le niveau de précision des simulations hydrodynamiques atteint celui des données, de telle sorte que nous devons maintenant prendre en compte les phénomènes de rétro-action des galaxies, liés aux supernovae et trous noirs supermassifs, appelés Active Galactic Nuclei (AGN). En effet, ces procesus de feedback rejettent une quantité considérable d'énergie dans le milieu environnant et modifient l'état thermique et la distribution du gaz dans le milieu intergalactique. Cette partie présente mon travail effectué à l'aide de simulations hydrodynamiques cosmologiques, que j'ai pour la grande majorité réalisée, pour casser la forte dégénérescence entre la cosmologie avec le $P_{\mathrm{Ly}\alpha}$ et les phénomènes de feedbacks, publié dans Chabanier et al. (2020).

B.3.1 La grille de simulations Horizon-AGN

Afin d'estimer l'impact du feedback des AGNs sur la forêt Lyα j'ai utilisé la simulation Horizon-AGN (HAGN) (Dubois et al., 2016) dont les caractéristiques en terme de résolution, taille de boîte et implémentation de processus physiques sont idéales pour l'étude des forêts Lyα. En particulier, elle a été réalisée avec le code à raffinement adaptatif de maille RAMSES (Teyssier, 2002) qui permet d'augmenter la résolution de la grille dans les régions de haute densité. On peut donc à la fois contrôler la résolution dans les régions de faibles densités, là où se situe le gaz Lyα, tout en simulant à haute résolution la physique des galaxies. De plus, inclure les phénomènes de feedback dans les simulations implique de calibrer des paramètres sous grille arbitraires sur des observations astrophysiques, ce qui conduit à de fortes tensions entre les prédictions de différentes simulations hydrodynamiques. HAGN a été calibrée à z = 0 sur la relation de Maggorian (Dubois et al., 2012; Volonteri et al., 2016) et reproduit bien la fraction de gaz dans les galaxies à différents redshifts, ce qui est l'un des grands problèmes des simulations hydrodynamiques cosmologiques et est un obvervable majeur pour l'étude des impacts des mécanismes de feedback sur la forêt Lyα.

Dans HAGN les trous noirs supermassifs sont implémentés avec des particules puits qui peuvent accreter du gaz au taux d'accrétion Bondi-Hoyle-Lyttleton

$$\dot{M}_{\mathrm{BH}} = \frac{4\pi\alpha G^2 M_{\mathrm{BH}}^2 \overline{\rho}}{(\overline{c}_s^2 + \overline{u}^2)^{3/2}}, \tag{B.12}$$

où α est le facteur "boost" sans dimension, M_{BH} la masse du trou noir, $\overline{\rho}$ la densité de gaz moyenne, $\overline{\mathrm{c}}_{\mathrm{s}}$ la vitesse du son moyenne et $\overline{\mathrm{u}}$ la vitesse relative moyenne du gaz par rapport au trou noir. $\alpha \geq 1$ permet de rendre compte du manque de résolution dans le disque d'accrétion. On a $\alpha = (\rho/\rho_0)^2$ si $\rho > \rho_0$ et $\alpha = 1$ sinon. $\dot{M}_{\mathrm{BH}}$ est limité par le taux d'accrétion Eddington,

$$\dot{M}_{\mathrm{Edd}} = \frac{4\pi G M_{\mathrm{BH}} m_{\mathrm{p}}}{\epsilon_{\mathrm{r}} \sigma_{\mathrm{T}} c}, \tag{B.13}$$

avec σ_{T} la section efficace Thomson, c la vitesse de la lumière et $\epsilon_{\mathrm{r}} = 0.1$ l'efficacité radiative. Le feedback des AGN injectent une fraction d'énergie ϵ_{f} de l'énergie lumineuse, $\mathrm{L_r}$, sous la forme d'énergie thermique et cinétique, tel que

$$\Delta \mathrm{E}_{\mathrm{medium}} = \epsilon_{\mathrm{f}} \mathrm{L_r} \tag{B.14}$$

$$= \epsilon_{\mathrm{f}} \epsilon_{\mathrm{r}} \dot{M}_{\mathrm{BH}}^2 c^2. \tag{B.15}$$

Simulation	α	$\mathrm{r}_{\mathrm{AGN}}$	ϵ_{f}
HAGN	$\begin{cases} (\rho/\rho_0)^2 & \text{si } \rho > \rho_0 \\ 1 & \text{sinon} \end{cases}$	Δx	$\begin{cases} 0.1 & \text{si } \chi < 10^{-2} \\ 0.15 & \text{si } \chi > 10^{-2} \end{cases}$
HAGNclp10	10% du temps : $10\alpha_{\mathrm{HAGN}}$	$\mathrm{r}_{\mathrm{AGN,HAGN}}$	$\epsilon_{\mathrm{f,HAGN}}$
HAGNclp100	1% du temps : $100\alpha_{\mathrm{HAGN}}$	$\mathrm{r}_{\mathrm{AGN,HAGN}}$	$\epsilon_{\mathrm{f,HAGN}}$
HAGNr+	α_{HAGN}	$2\Delta x$	$\epsilon_{\mathrm{f,HAGN}}$
HAGNr−	α_{HAGN}	$0.5\Delta x$	$\epsilon_{\mathrm{f,HAGN}}$
HAGNϵ+	α_{HAGN}	$\mathrm{r}_{\mathrm{AGN,HAGN}}$	$\begin{cases} 3 & \text{si } \chi < 10^{-2} \\ 0.45 & \text{si } \chi > 10^{-2} \end{cases}$
HAGNϵ−	α_{HAGN}	$\mathrm{r}_{\mathrm{AGN,HAGN}}$	$\begin{cases} 0.33 & \text{si } \chi < 10^{-2} \\ 0.05 & \text{si } \chi > 10^{-2} \end{cases}$

Table B.1 – *Résumé des paramètres sous grille pour la série de simulations réalisées pour explorer une grande gamme de modèles de feedback plausibles. De gauche à droite les colonnes donnent : le nom de la simulation, la valeur du facteur boost* α*, the rayon de depôt d'énergie* r_{agn} *où* Δx *est la taille de la plus petite cellule, et finalement la fraction d'énergie lumineuse injectée dans le milieu environnant* ϵ_{f}.

et dans un rayon $\mathrm{r}_{\mathrm{AGN}}$. Le feedback est implémenté en deux modes (Dubois et al., 2012) selon la valeur du rapport du taux d'accrétion à la limite d'Eddington

$$\chi = \frac{\dot{M}_{\mathrm{BH}}}{\dot{M}_{\mathrm{Edd}}}. \tag{B.16}$$

Si $\chi > 10^{-2}$ le mode quasar est déclenché et injecte de l'énergie thermique. Sinon le mode radio est déclenché et injecte de l'énergie cinétique. Une simulation compagnon, Horizon-noAGN (HnoAGN), a été tournée avec les mêmes caractéristiques, conditions initiales et modèles physiques mais sans feedback des AGN.

Afin d'estimer les incertitudes liées aux différentes paramétrisations et calibrations des mécanismes de feedback, j'ai construit une grille de six simulations qui vise à explorer une grande gamme de modèles de feedback plausibles en modifiant les paramètres sous grille principaux susceptibles d'impacter l'évolution de la forêt Lyα :

- **HAGNclp10** et **HAGNclp100** introduisent de la stochasticité dans le taux d'accrétion.
- **HAGNr**+ et **HAGNr**− modifient le rayon de depôt d'énergie, $\mathrm{r}_{\mathrm{AGN}}$.
- **HAGN**ϵ+ et **HAGN**ϵ− modifient la fraction d'énergie injectée dans le milieu ϵ_{f}.

Les valeurs des paramètres sous grille pour chacune des simulations sont données dans la Table B.1. Elles ont été choisies de tel sorte qu'on balaye les incertitudes observationnelles liées à la fraction de gaz moyenne dans les galaxies, $\mathrm{f}_{\mathrm{gas}}$, et la relation de Maggorian, $\mathrm{M}_{\mathrm{BH}} - \mathrm{M}_{*}$. La table B.2 montre qu'on couvre à plus de 3σ ces observables astrophysiques ce qui assure une exploration des modèles de feedback à la limite du plausible.

B.3.2 Impact des phénomènes de feedbacks

Figure B.7 montre l'impact du feedback des AGN sur $P_{\mathrm{Ly}\alpha}$en utilisant les simulations HAGN et HnoAGN, ainsi que l'incertitude liée aux différents modèles de feedback estimée avec la grille de simulations. On observe une suppression de puissance qui augmente avec l'échelle et lorsque le redshift décroit. La suppression de puissance atteint plus de 10% à $z = 2$ et aux grandes échelles, soit bien supérieur au niveau de sensibilité des données qui est proche du pourcent à

	$\Delta\sigma_{\mathrm{f_{gas}}}$	$\Delta\sigma_{\mathrm{M_{BH}-M_*}}$
HAGN	0	0
HAGNclp10	$< \sigma_{\mathrm{f_{gas}}}$	$\sigma_{\mathrm{M_{BH}-M_*}}$
HAGNclp100	$\sigma_{\mathrm{f_{gas}}}$	$\sigma_{\mathrm{M_{BH}-M_*}}$
HAGNr+	$3\sigma_{\mathrm{f_{gas}}}$	$2\sigma_{\mathrm{M_{BH}-M_*}}$
HAGNr−	$2.7\sigma_{\mathrm{f_{gas}}}$	$3.3\sigma_{\mathrm{M_{BH}-M_*}}$
HAGNϵ+	$2.3\sigma_{\mathrm{f_{gas}}}$	$3.5\sigma_{\mathrm{M_{BH}-M_*}}$
HAGNϵ−	$2.5\sigma_{\mathrm{f_{gas}}}$	$3.5\sigma_{\mathrm{M_{BH}-M_*}}$

Table B.2 – *Déviations d'observables astrophysiques entre* HAGN *et les autres simulations, exprimées en terme d'incertitude observationnelle à $z = 2$. Les observables astrophysiques étudiées sont la fraction de gaz moyenne dans les galaxies (au milieu) et la relation de Maggorian entre la masse du trou noir,* $\mathrm{M_{BH}}$*, et la masse stellaire des galaxies,* $\mathrm{M_*}$ *(à droite).*

ce redshift, ce qui justifie la nécessité de prendre en compte ce phénomène dans nos prédictions théoriques.

Sachant que le feedback des AGNs rejette du gaz depuis les petites échelles galactiques aux grandes échelles du milieu intergalactique, nous aurions pu naïvement nous attendre à une augmentation de puissance aux grandes échelles, là où le gaz est retombé. Nous expliquons cette suppression de puissance par un chauffage très efficace du gaz Lyα. En effet, le milieu intergalactique chaud contient 18% de la masse totale du gaz dans HAGN et seulement 12% dans HnoAGN. On peut donc en déduire que le feedback des AGNs produit un chauffage très important du milieu intergalactique qui conduit à sa ionization et donc à une suppression de puissance étant donné que $P_{\mathrm{Ly}\alpha}$ sonde l'hydrogène neutre et non l'hydrogène ionisé.

B.3.3 Résolution dans le milieu intergalactique : la simulation Extreme-Horizon

De la section précédente, nous pouvons conclure que les effets thermiques sont dominants par rapport aux effets de redistribution du gaz. Néanmoins, les processus de chauffage et de refroidissement sont mal maitrisés dans les simulations numériques et sont très dépendents de la résolution. Afin de tester des effets systématiques sur la correction de feedback AGN liés à la résolution du milieu intergalactique, j'ai initié et réalisé la simulaiton Extreme-Horizon (EH) dans le cadre d'un "Grand Challenge", présentée en détail dans Chabanier et al. (2020). Par rapport à HAGN, EH double la résolution de la grille partout sauf dans les galaxies qui sont traitées à la même résolution, et dans un volume plus petit. J'ai également réalisé la simulation Standard-Horizon (SH) dans le même volume que EH et avec la même stratégie en résolution que dans HAGN.

La comparaison des $P_{\mathrm{Ly}\alpha}$ de EH et SH montre un effet très large, qui atteint 20% aux plus petites échelles ($k \sim 0.025$ s/km) indiquant que celui ci n'est pas convergé en absolu dans SH, et donc par conséquent dans HAGN. Cependant, les corrections de feedback AGN de EH et HAGN présentent des différences bien inférieures au pourcent, ce qui signifie que cette dernière est bien convergée dans HAGN et qu'il n'y a donc pas de couplage entre la résolution des grandes échelles du milieu intergalactique et la correction de feedback AGN.

Finalement, durant les derniers mois de ma thèse, j'ai réalisé des simulations similaires visant à étudier l'impact du feedback des supernovae sur la forêt Lyα qui montrent que leur impact seul

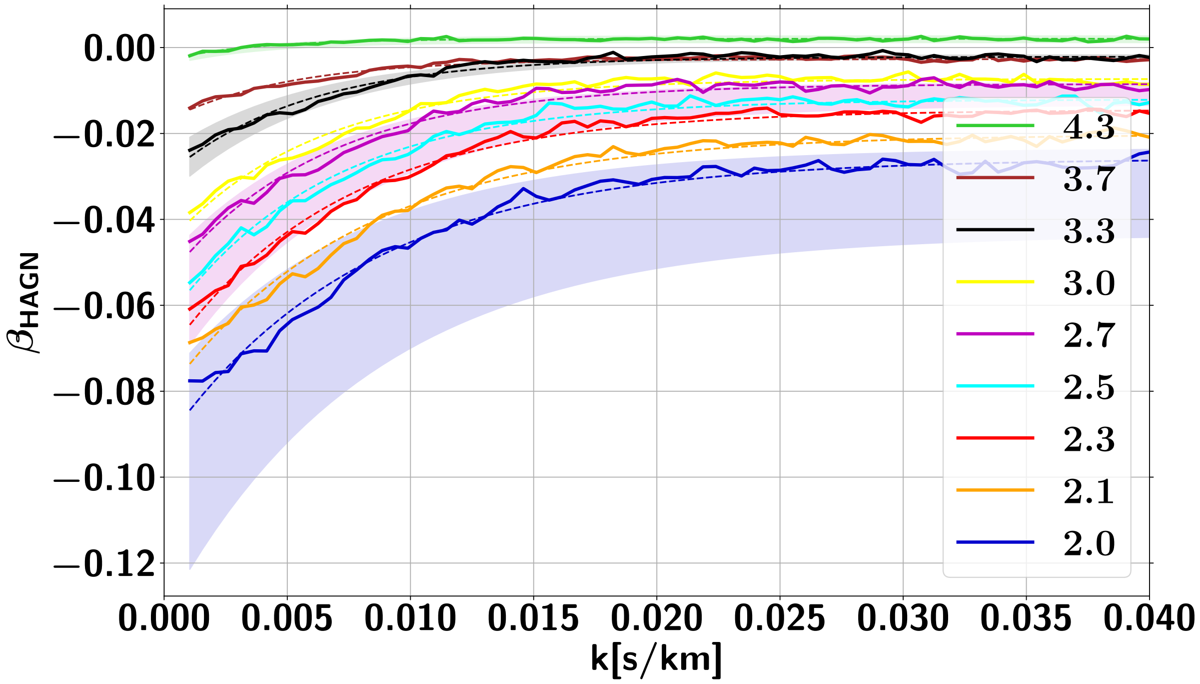

Figure B.7 – *Corrections dues aux feedback baryoniques sur la forêt Lyα à tous les redshifts. Les incertitudes sont montrées par les bandes colorées, seulement aux redshifts $z = 4.25$, $z = 2.7$ et $z = 2$ pour une meilleure lisibilité. Les incertitudes sont seulement systématiques, léies aux icnertitudes de modélisation et de calibration du feedback.*

est négligeable étant donné la précision actuelle des données. Nous pouvons donc en conclure que la correction précédemment établie englobe tous les mécanismes de feedback baryoniques ainsi que leur couplage. Nous montrons aussi qu'ignorer les feedbacks baryoniques dans les prédictions théoriques de $P_{\mathrm{Ly}\alpha}$ conduit à un biais à 2σ sur n_s et à 1σ sur σ_8 en utilisant les données eBOSS-Lyα. Etant donné que les incertitudes observationelles vont être drastiquement réduites avec l'expérience nouvelle génération DESI, ces biais systématiques seront encore plus importants, ce qui justifie encore une fois la nécessité de casser la dégénérescence entre la cosmologie et la physique baryonique.

B.4 Résultats cosmologiques

Les mesures de spectre de puissance des forêts Lyα effectuées à partir des relevés spectroscopiques BOSS et eBOSS peuvent être confrontées statistiquement aux prédictions théoriques issues de simulations numériques qui incluent toute la physique nécessaire à la bonne modélisation de la forêt Lyα pour inverser les données et contraindre la cosmologie. Je présente dans cette partie les résultats cosmologiques principaux, dont l'étude a été publiée dans Palanque-Delabrouille et al. (2020).

	Lyα	P18+Lyα	Lyα+P18 +lens. +BAO
T_0 (z=3) (10^3K)...	8.5 ± 2.0	7.8 ± 1.8	8.0 ± 1.8
γ	0.93 ± 0.14	0.80 ± 0.12	0.80 ± 0.12
σ_8	0.826 ± 0.021	0.825 ± 0.007	0.821 ± 0.007
n_s	0.954 ± 0.006	0.962 ± 0.003	0.962 ± 0.003
Ω_m	0.269 ± 0.009	0.306 ± 0.007	0.307 ± 0.006
α_s	-	-0.010 ± 0.003	-0.010 ± 0.003
$\sum m_\nu$ (eV , 95% CL)	< 0.58	< 0.105	< 0.089

Table B.3 – *Paramètres astrophysiques et cosmologiques estimées à un niveau de confiance de 68.3% pour le modèle ΛCDMν pour les données Lyα (à gauche), et pour le modèle ΛCDMν+α_s pour la combinaison des données Lyα et CMB (au milieu) et la combinaison des données Lyα, CMB et BAO (à droite).*

B.4.1 Masse des neutrinos

Les contraintes sur les paramètres cosmologiques et astrophysiques, notamment sur $\sum m_\nu$, en combinant différents jeux de données sont résumées dans la table B.3. Au total nous utilisons 22 paramètres :

- 6 paramètres cosmologiques dans l'espace $(\Omega_m, H_0, \sigma_8, n_s, \sum m_\nu, \alpha_s)$. Le paramètre $\alpha_s \equiv dn_s/d\ln k$ permet d'introduire une dépendance d'échelle pour n_s afin de combiner les données Lyα et CMB-Planck 2018 qui présentent une tension à 3.6σ sur Ω_m qui peut également être intepretée comme une tension sur n_s.
- 10 paramètres astrophysiques liés à la modélisation de la température dans le milieu intergalactique, la relation température densité du gaz du milieu intergalactique et aux absorptions corrélées aux absorption Lyα, i.e. Si III et Si II.
- 6 paramètres de nuisance qui rendent compte des incertitudes de mesures et de modelisation dans les simulations numériques.

Dans un premier temps, en ne prenant en compte que les données eBOSS-Lyα, nous contraingons $\sum m_\nu < 0.58$ eV, soit une contrainte deux fois meilleure à celle publiée dans Palanque-Delabrouille et al. (2015) qui utilise les données DR9 BOSS-Lyα. Cette amélioration d'un facteur deux est dûe à la précision sans précédent de la nouvelle analyse de donnée, grâce à un plus grand jeu de données et une analyse plus fine et rigoureuse des incertitude systématiques.

Cependant, la forêt Lyα sonde une gamme d'échelle où la suppression de la formation des structures causée par les neutrinos massifs est relativement indépendante de l'échelle, ce qui conduit à une forte dégénérescence, de l'ordre de 70%, entre $\sum m_\nu$ et σ_8, l'amplitude du spectre de puissance. Les données CMB (Planck Collaboration et al., 2018b) sondent des échelles où la formation des structures n'est pas impactée par la présence de neutrinos massifs, et donc la combinaison de ces deux jeux de données permet de casser la dégénérescence entre amplitude et masse de neutrinos et ainsi renforcer la contrainte à $\sum m_\nu < 0.105$ eV.

Finalement, ajouter les données sondant les échelles BAO tel 6dFGS (Beutler et al., 2011), SDSS main galaxy smple (Ross et al., 2014) et CMASS-DR12 (Alam et al., 2017) renforce encore la contrainte à $\sum m_\nu < 0.089$ eV, ce qui est l'une des contraintes les plus fortes à ce jour et semble favoriser un scénario de masse dit normal.

	m_x (keV , 95% CL)	m_{ν_s} (keV , 95% CL)
Lyα+XQ-100	8.6	63
Lyα(z<4.5)+XQ-100	5.3	34
Lyα(z<4.1)+XQ-100	3.9	22

TABLE B.4 – *Contraintes sur la masse de reliques thermiques et de neutrinos stériles non résonants (niveau de confiance à 68.3%) dans un modèle de matière noire composé entièrement de matière noire tiède en combinant les donnés eBOSS-Lyα and XQ-100 pour différentes sélection en redshifts pour les données eBOSS-Lyα.*

B.4.2 Matière noire tiède

Pour contraindre la masse des reliques thermiques, m_x, on utilise une combinaison des données eBOSS-Lyα avec 100 spectres très haute résolution du relevé XQ-100 analysés dans Yèche et al. (2017). La résolution trois à cinq fois meilleure permet de sonder des échelles plus petites que celles sondées par SDSS, là où l'impact de la matière noire tiède est plus plus importante. Nous nous plaçons dans un modèle ΛWDM. Les paramètres cosmologiques considérés ici sont $(\Omega_m, H_0, \sigma_8, n_s, 1/m_x)$. Comme le modèle ΛCDM est reproduit pour $m_x \to \infty$, nous considérons $1/m_x$ au lieu de m_x.

La combinaison des données contraint la masse d'un relique thermique à $m_x > 8.6$ keV (95% CL). Puisque la croissance non-linéaire des structures tend à alléger la suppression causée par la libre propagation des particules relativistes, les $P_{\mathrm{Ly}\alpha}$ à hauts redshifts sont plus sensibles à m_x. Cependant ces redshifts sont aussi les plus sensibles aux systématiques (voir figure B.5), Tab. B.4 présente donc les contraintes sur m_x en considérant différentes sélections sur les bins en redshifts. Nous considérons comme résultat le plus conservatif celui obtenu dans la configuration eBOSS-Lyα $(z < 4.5)$ + XQ-100 donnant $m_x > 5.3$ keV (95% CL). Ce résultat est l'un des plus forts à ce jour et significativement plus contraignant que celui obtenu avec BOSS-Lyα $(z < 4.5)$ + XQ-100 donnant $m_x > 4.2$ keV (95% CL) (Yèche et al., 2017).

Bibliography

Abbott, B. P., R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, et al. (2017). A gravitational-wave standard siren measurement of the Hubble constant. Nature 551(7678), 85–88. arXiv:astro-ph.CO/1710.05835.

Abbott, T. M. C., F. B. Abdalla, S. Allam, A. Amara, J. Annis, et al. (2018). The Dark Energy Survey: Data Release 1. Astrophysical Journal Supplement Series 239(2), 18. arXiv:astro-ph.IM/1801.03181.

Abe, K., R. Akutsu, A. Ali, C. Alt, C. Andreopoulos, et al. (2019). Constraint on the Matter-Antimatter Symmetry-Violating Phase in Neutrino Oscillations. arXiv e-prints, arXiv:1910.03887. arXiv:hep-ex/1910.03887.

Abel, T., P. Anninos, Y. Zhang, and M. L. Norman (1997). Modeling primordial gas in numerical cosmology. New Astron. 2, 181–207. astro-ph/9608040.

Abolfathi, B., D. S. Aguado, G. Aguilar, C. Allende Prieto, A. Almeida, et al. (2018). The Fourteenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the Extended Baryon Oscillation Spectroscopic Survey and from the Second Phase of the Apache Point Observatory Galactic Evolution Experiment. Astrophysical Journal Supplement Series 235, 42. arXiv:1707.09322.

Addison, G. E., D. J. Watts, C. L. Bennett, M. Halpern, G. Hinshaw, et al. (2018). Elucidating ΛCDM: Impact of Baryon Acoustic Oscillation Measurements on the Hubble Constant Discrepancy. Astrophysical Journal 853(2), 119. arXiv:astro-ph.CO/1707.06547.

Ade, P. A. R. et al. (2014). Planck 2013 results. XVI. Cosmological parameters. Astron. Astrophys. 571, A16. arXiv:astro-ph.CO/1303.5076.

Agertz, O., A. V. Kravtsov, S. N. Leitner, and N. Y. Gnedin (2013). Toward a Complete Accounting of Energy and Momentum from Stellar Feedback in Galaxy Formation Simulations. Astrophysical Journal 770(1), 25. arXiv:astro-ph.CO/1210.4957.

Agertz, O., B. Moore, J. Stadel, D. Potter, F. Miniati, et al. (2007). Fundamental differences between SPH and grid methods. MNRAS 380(3), 963–978. arXiv:astro-ph/astro-ph/0610051.

Agertz, O., R. Teyssier, and B. Moore (2011). The formation of disc galaxies in a ΛCDM universe. MNRAS 410(2), 1391–1408. arXiv:astro-ph.CO/1004.0005.

Aghanim, N. et al. (2019). Planck 2018 results. V. CMB power spectra and likelihoods. arXiv:astro-ph.CO/1907.12875.

Ahmad, Q. R., R. C. Allen, T. C. Andersen, J. D. Anglin, G. Bühler, et al. (2001). Measurement of the Rate of ν_e + d –> p + p + e^- Interactions Produced by ^{8}B Solar Neutrinos at the Sudbury Neutrino Observatory. Phys. Rev. Lett. 87(7), 071301. arXiv:nucl-ex/nucl-ex/0106015.

Ahn, C. P., Alexandroff, J. P. Wisniewski, W. M. Wood-Vasey, B. Yanny, et al. (2012). The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey. ApJS 203(2), 21. arXiv:1207.7137.

Ahumada, R., C. Allende Prieto, A. Almeida, F. Anders, S. F. Anderson, et al. (2019). 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 e-prints, arXiv:1912.02905. arXiv:astro-ph.GA/1912.02905.

Aihara, H., C. Allende Prieto, D. An, S. F. Anderson, É. Aubourg, et al. (2011). The Eighth Data Release of the Sloan Digital Sky Survey: First Data from SDSS-III. Astrophysical Journal Supplement Series 193(2), 29. arXiv:astro-ph.IM/1101.1559.

Aker, M., K. Altenmüller, M. Arenz, M. Babutzka, J. Barrett, et al. (2019). Improved Upper Limit on the Neutrino Mass from a Direct Kinematic Method by KATRIN. Phys. Rev. Lett. 123(22), 221802. arXiv:hep-ex/1909.06048.

Alam, S., M. Ata, S. Bailey, F. Beutler, D. Bizyaev, et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. MNRAS 470(3), 2617–2652. arXiv:astro-ph.CO/1607.03155.

Alard, C. (2014). The Baryonic Self Similarity of Dark Matter. Astrophysical Journal 788(2), 171. arXiv:astro-ph.CO/1305.5738.

Ali-Haïmoud, Y. and S. Bird (2013). An efficient implementation of massive neutrinos in non-linear structure formation simulations. MNRAS 428(4), 3375–3389. arXiv:astro-ph.CO/1209.0461.

Aprile, E., J. Aalbers, F. Agostini, M. Alfonsi, L. Althueser, et al. (2020). Observation of Excess Electronic Recoil Events in XENON1T. arXiv e-prints, arXiv:2006.09721. arXiv:hep-ex/2006.09721.

Archidiacono, M. and S. Hannestad (2016). Efficient calculation of cosmological neutrino clustering in the non-linear regime. J. Cosmology Astropart. Phys. 2016(6), 018. arXiv:astro-ph.CO/1510.02907.

Archidiacono, M., S. Hannestad, and J. Lesgourgues (2020). What will it take to measure individual neutrino mass states using cosmology? arXiv e-prints, arXiv:2003.03354. arXiv:astro-ph.CO/2003.03354.

Armengaud, E., N. Palanque-Delabrouille, C. Yèche, D. J. E. Marsh, and J. Baur (2017). Constraining the mass of light bosonic dark matter using sdss lyman-α forest. Monthly Notices of the Royal Astronomical Society 471(4), 4606–4614.

Asaka, T., S. Blanchet, and M. Shaposhnikov (2005). The νMSM, dark matter and neutrino masses [rapid communication]. Physics Letters B 631(4), 151–156. arXiv:astro-ph/hep-ph/0503065.

Asaka, T., M. Shaposhnikov, and A. Kusenko (2006). Opening a new window for warm dark matter. Physics Letters B 638(5-6), 401–406. arXiv:hep-ph/hep-ph/0602150.

Asgari, M. et al. (2019). KiDS+VIKING-450 and DES-Y1 combined: Mitigating baryon feedback uncertainty with COSEBIs. arXiv:astro-ph.CO/1910.05336.

Ata, M., F. Baumgarten, J. Bautista, F. Beutler, D. Bizyaev, et al. (2018). 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. MNRAS 473, 4773–4794. arXiv:1705.06373.

Aubert, D., C. Pichon, and S. Colombi (2004). The origin and implications of dark matter anisotropic cosmic infall on $\sim L_*$ haloes. MNRAS 352(2), 376–398. arXiv:astro-ph/astro-ph/0402405.

Banerjee, A. and N. Dalal (2016). Simulating nonlinear cosmological structure formation with massive neutrinos. J. Cosmology Astropart. Phys. 2016(11), 015. arXiv:astro-ph.CO/1606.06167.

Banerjee, A., D. Powell, T. Abel, and F. Villaescusa-Navarro (2018). Reducing noise in cosmological N-body simulations with neutrinos. J. Cosmology Astropart. Phys. 2018(9), 028. arXiv:astro-ph.CO/1801.03906.

Bardos, C., I. M. Gamba, F. Golse, and C. D. Levermore (2016). Global Solutions of the Boltzmann Equation Over {{R}$\hat{\mathrm{D}}$} Near Global Maxwellians with Small Mass. Communications in Mathematical Physics 346(2), 435–467. arXiv:math.AP/1409.1430.

Barnes, J. and P. Hut (1986). A hierarchical O(N log N) force-calculation algorithm. Nature 324(6096), 446–449.

Baron, D. and B. Ménard (2019). Black hole mass estimation for active galactic nuclei from a new angle. MNRAS 487(3), 3404–3418. arXiv:astro-ph.GA/1903.01996.

Barro, G., S. M. Faber, D. C. Koo, A. Dekel, J. J. Fang, et al. (2017). Structural and Star-forming Relations since $z = 3$: Connecting Compact Star-forming and Quiescent Galaxies. Astrophysical Journal 840(1), 47. arXiv:astro-ph.GA/1509.00469.

Barro, G., S. M. Faber, P. G. Pérez-González, D. C. Koo, C. C. Williams, et al. (2013). CANDELS: The Progenitors of Compact Quiescent Galaxies at z $\sim$2. Astrophysical Journal 765(2), 104. arXiv:astro-ph.CO/1206.5000.

Baugh, C. M. (2006). A primer on hierarchical galaxy formation: the semi-analytical approach. Reports on Progress in Physics 69(12), 3101–3156.

Baugh, C. M., E. Gaztanaga, and G. Efstathiou (1995). A comparison of the evolution of density fields in perturbation theory and numerical simulations - II. Counts-in-cells analysis. MNRAS 274(4), 1049–1070. arXiv:astro-ph/astro-ph/9408057.

Baumann, D., F. Beutler, R. Flauger, D. Green, A. Slosar, et al. (2019). First constraint on the neutrino-induced phase shift in the spectrum of baryon acoustic oscillations. Nature Physics 15(5), 465–469. arXiv:astro-ph.CO/1803.10741.

Baur, J., N. Palanque-Delabrouille, C. Yèche, A. Boyarsky, O. Ruchayskiy, et al. (2017). Constraints from ly-alpha forests on non-thermal dark matter including resonantly-produced sterile neutrinos. Journal of Cosmology and Astroparticle Physics 2017(12), 013–013.

Baur, J., N. Palanque-Delabrouille, C. Yèche, C. Magneville, and M. Viel (2016). Lyman-alpha forests cool warm dark matter. Journal of Cosmology and Astroparticle Physics 2016(08), 012–012.

Bautista, J. E., N. G. Busca, J. Guy, J. Rich, M. Blomqvist, et al. (2017). Measurement of baryon acoustic oscillation correlations at $z = 2.3$ with sdss dr12 lya-forests. Astronomy & Astrophysics 603, A12.

Bechtold, J., A. P. S. Crotts, R. C. Duncan, and Y. Fang (1994). Spectroscopy of the Double Quasars Q1343+266A, B: A New Determination of the Size of Lyman-Alpha Forest Absorbers. Astrophysical Journal Letters 437, L83. arXiv:astro-ph/astro-ph/9409007.

Beck, A. M., G. Murante, A. Arth, R. S. Remus, A. F. Teklu, et al. (2016). An improved SPH scheme for cosmological simulations. MNRAS 455(2), 2110–2130. arXiv:astro-ph.CO/1502.07358.

Becker, G. D., J. S. Bolton, M. G. Haehnelt, and W. L. W. Sargent (2011). Detection of extended HeII reionization in the temperature evolution of the intergalactic medium. Monthly Notices of the Royal Astronomical Society 410(2), 1096–1112. arXiv:1008.2622.

Becker, G. D., W. L. W. Sargent, and M. Rauch (2004). Large-Scale Correlations in the Lyα Forest at z = 3-4. Astrophysical Journal 613(1), 61–76. arXiv:astro-ph/astro-ph/0405205.

Beckmann, R. S., J. Devriendt, A. Slyz, S. Peirani, M. L. A. Richardson, et al. (2017). Cosmic evolution of stellar quenching by AGN feedback: clues from the Horizon-AGN simulation. MNRAS 472(1), 949–965. arXiv:astro-ph.GA/1701.07838.

Begeman, K. G., A. H. Broeils, and R. H. Sanders (1991). Extended rotation curves of spiral galaxies : dark haloes and modified dynamics. MNRAS 249, 523.

Bennett, C. L., D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, et al. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. Astrophysical Journal Supplement Series 208(2), 20. arXiv:astro-ph.CO/1212.5225.

Benson, A. J., R. G. Bower, C. S. Frenk, C. G. Lacey, C. M. Baugh, et al. (2003). What Shapes the Luminosity Function of Galaxies? Astrophysical Journal 599(1), 38–49. arXiv:astro-ph/astro-ph/0302450.

Berger, M. J. and P. Colella (1989). Local Adaptive Mesh Refinement for Shock Hydrodynamics. Journal of Computational Physics 82(1), 64–84.

Berger, M. J. and J. Oliger (1984). Adaptive Mesh Refinement for Hyperbolic Partial Differential Equations. Journal of Computational Physics 53(3), 484–512.

Bertschinger, E. (2001). Multiscale Gaussian Random Fields and Their Application to Cosmological Simulations. Astrophysical Journal Supplement Series 137(1), 1–20. arXiv:astro-ph/astro-ph/0103301.

Beutler, F., C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, et al. (2011). The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant. MNRAS 416, 3017–3032. arXiv:astro-ph.CO/1106.3366.

Bezrukov, F., H. Hettmansperger, and M. Lindner (2010). keV sterile neutrino dark matter in gauge extensions of the standard model. Phys. Rev. D 81(8), 085032. arXiv:hep-ph/0912.4415.

Bi, H. and A. F. Davidsen (1997). Evolution of Structure in the Intergalactic Medium and the Nature of the Lyα Forest. Astrophysical Journal 479(2), 523–542. arXiv:astro-ph/astro-ph/9611062.

Biernacki, P. and R. Teyssier (2018). The combined effect of AGN and supernovae feedback in launching massive molecular outflows in high-redshift galaxies. MNRAS 475(4), 5688–5703. arXiv:astro-ph.GA/1712.02794.

Bird, S., M. Viel, and M. G. Haehnelt (2012). Massive neutrinos and the non-linear matter power spectrum. MNRAS 420(3), 2551–2561. arXiv:astro-ph.CO/1109.4416.

Blanton, M. R., M. A. Bershady, B. Abolfathi, F. D. Albareti, C. Allende Prieto, et al. (2017). Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe. Astronomical Journal 154, 28. arXiv:1703.00052.

Bode, P., J. P. Ostriker, and N. Turok (2001). Halo Formation in Warm Dark Matter Models. Astrophysical Journal 556(1), 93–107. arXiv:astro-ph/astro-ph/0010389.

Bolton, A. S., D. J. Schlegel, E. Aubourg, S. Bailey, V. Bhardwaj, et al. (2012). Spectral Classification and Redshift Measurement for the SDSS-III Baryon Oscillation Spectroscopic Survey. The Astronomical Journal 144(5), 144. arXiv:1207.7326.

Bolton, J. S. and G. D. Becker (2009). Resolving the high redshift lyα forest in smoothed particle hydrodynamics simulations. Monthly Notices of the Royal Astronomical Society: Letters 398(1), L26–L30. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1745-3933.2009.00700.x.

Bolton, J. S., G. D. Becker, M. G. Haehnelt, and M. Viel (2014). A consistent determination of the temperature of the intergalactic medium at redshift z = 2.4. MNRAS 438(3), 2499–2507. arXiv:astro-ph.CO/1308.4411.

Bond, J. R., G. Efstathiou, and J. Silk (1980). Massive Neutrinos and the Large-Scale Structure of the Universe. Phys. Rev. Lett. 45(24), 1980–1984.

Bond, J. R., A. S. Szalay, and J. Silk (1988). Lyman-Alpha Clouds as a Relic of Primordial Density Fluctuations. Astrophysical Journal 324, 627.

Booth, C. M. and J. Schaye (2009). Simulations of the Growth of Black Holes and Feedback from Active Galactic Nuclei. In S. Heinz and E. Wilcots (Eds.), American Institute of Physics Conference Series, Volume 1201 of American Institute of Physics Conference Series, pp. 21–24.

Booth, C. M. and J. Schaye (2011). Towards an understanding of the evolution of the scaling relations for supermassive black holes. MNRAS 413(2), 1158–1164. arXiv:astro-ph.CO/1005.0844.

Borde, A., N. Palanque-Delabrouille, G. Rossi, M. Viel, J. S. Bolton, et al. (2014). New approach for precise computation of Lyman-α forest power spectrum with hydrodynamical simulations. J. Cosmology Astropart. Phys. 2014(7), 005. arXiv:astro-ph.CO/1401.6472.

Bournaud, F., B. G. Elmegreen, and D. M. Elmegreen (2007). Rapid Formation of Exponential Disks and Bulges at High Redshift from the Dynamical Evolution of Clump-Cluster and Chain Galaxies. Astrophysical Journal 670(1), 237–248. arXiv:astro-ph/0708.0306.

Bournaud, F., C. J. Jog, and F. Combes (2007). Multiple minor mergers: formation of elliptical galaxies and constraints for the growth of spiral disks. Astronomy & Astrophysics 476(3), 1179–1190. arXiv:astro-ph/0709.3439.

Boyarsky, A., J. Lesgourgues, O. Ruchayskiy, and M. Viel (2009). Realistic Sterile Neutrino Dark Matter with KeV Mass does not Contradict Cosmological Bounds. Phys. Rev. Lett. 102(20), 201304. arXiv:hep-ph/0812.3256.

Boyarsky, A., D. Malyshev, O. Ruchayskiy, and D. Savchenko (2020). Technical comment on the paper of Dessert et al. “The dark matter interpretation of the 3.5 keV line is inconsistent with blank-sky observations”. arXiv e-prints, arXiv:2004.06601. arXiv:astro-ph.CO/2004.06601.

Boyarsky, A., O. Ruchayskiy, D. Iakubovskyi, and J. Franse (2014). Unidentified Line in X-Ray Spectra of the Andromeda Galaxy and Perseus Galaxy Cluster. Phys. Rev. Lett. 113(25), 251301. arXiv:astro-ph.CO/1402.4119.

Boyle, B. J. and R. J. Terlevich (1998). The cosmological evolution of the QSO luminosity density and of the star formation rate. MNRAS 293(2), L49–L51. arXiv:astro-ph/astro-ph/9710134.

Bozek, B., M. Boylan-Kolchin, S. Horiuchi, S. Garrison-Kimmel, K. Abazajian, et al. (2016). Resonant sterile neutrino dark matter in the local and high-z Universe. MNRAS 459(2), 1489–1504. arXiv:astro-ph.CO/1512.04544.

Brandbyge, J. and S. Hannestad (2009). Grid based linear neutrino perturbations in cosmological N-body simulations. J. Cosmology Astropart. Phys. 2009(5), 002. arXiv:astro-ph/0812.3149.

Brandbyge, J. and S. Hannestad (2010). Resolving cosmic neutrino structure: a hybrid neutrino N-body scheme. J. Cosmology Astropart. Phys. 2010(1), 021. arXiv:astro-ph.CO/0908.1969.

Bressand, O., L. Colombet, A. Fontaine, G. Harel, and J.-B. Lekien (2012). in CHOCS, 41, 29 (http://www-physique-chimie.cea.fr/science-en-ligne/docs/chocs/Chocs_41.pdf).

Bryan, G. L., M. L. Norman, B. W. O’Shea, T. Abel, J. H. Wise, et al. (2014). ENZO: An Adaptive Mesh Refinement Code for Astrophysics. Astrophysical Journal Supplement Series 211(2), 19. arXiv:astro-ph.IM/1307.2265.

Buen-Abad, M. A., G. Marques-Tavares, and M. Schmaltz (2015). Non-Abelian dark matter and dark radiation. Phys. Rev. D92(2), 023531. `arXiv:hep-ph/1505.03542`.

Bulbul, E., M. Markevitch, A. Foster, R. a. K. Smith, M. Loewenstein, et al. (2014). Detection of an Unidentified Emission Line in the Stacked X-Ray Spectrum of Galaxy Clusters. Astrophysical Journal 789(1), 13. `arXiv:astro-ph.CO/1402.2301`.

Burles, S., K. M. Nollett, and M. S. Turner (2001). Big Bang Nucleosynthesis Predictions for Precision Cosmology. Astrophysical Journal Letters 552(1), L1–L5. `arXiv:astro-ph/astro-ph/0010171`.

Busca, N. G., T. Delubac, J. Rich, S. Bailey, A. Font-Ribera, et al. (2013). Baryon acoustic oscillations in the Ly-alpha forest of BOSS quasars. Astronomy & Astrophysics 552, A96. `arXiv:1211.2616`.

Capozziello, S. and G. Lambiase (2002). Brans-Dicke Theory of Gravity and Flavour Oscillations. In V. G. Gurzadyan, R. T. Jantzen, and R. Ruffini (Eds.), The Ninth Marcel Grossmann Meeting, pp. 2407–2408.

Cappelluti, N., Y. Li, A. Ricarte, B. Agarwal, V. Allevato, et al. (2017). The Chandra COSMOS Legacy Survey: Energy Spectrum of the Cosmic X-Ray Background and Constraints on Undetected Populations. Astrophysical Journal 837(1), 19. `arXiv:astro-ph.HE/1702.01660`.

Carbone, C., M. Petkova, and K. Dolag (2016). DEMNUni: ISW, Rees-Sciama, and weak-lensing in the presence of massive neutrinos. J. Cosmology Astropart. Phys. 2016(7), 034. `arXiv:astro-ph.CO/1605.02024`.

Carollo, C. M., T. J. Bschorr, A. Renzini, S. J. Lilly, P. Capak, et al. (2013). Newly Quenched Galaxies as the Cause for the Apparent Evolution in Average Size of the Population. Astrophysical Journal 773(2), 112. `arXiv:astro-ph.CO/1302.5115`.

Carroll, S. M. (2004). Why is the Universe Accelerating? In R. E. Allen, D. V. Nanopoulos, and C. N. Pope (Eds.), The New Cosmology: Conference on Strings and Cosmology, Volume 743 of American Institute of Physics Conference Series, pp. 16–32.

Castorina, E., E. Sefusatti, R. K. Sheth, F. Villaescusa-Navarro, and M. Viel (2014). Cosmology with massive neutrinos II: on the universality of the halo mass function and bias. J. Cosmology Astropart. Phys. 2014(2), 049. `arXiv:astro-ph.CO/1311.1212`.

Cen, R., J. Miralda-Escudé, J. P. Ostriker, and M. Rauch (1994). Gravitational Collapse of Small-Scale Structure as the Origin of the Lyman-Alpha Forest. Astrophysical Journal Letters 437, L9. `arXiv:astro-ph/astro-ph/9409017`.

Ceverino, D., A. Klypin, E. S. Klimek, S. Trujillo-Gomez, C. W. Churchill, et al. (2014). Radiative feedback and the low efficiency of galaxy formation in low-mass haloes at high redshift. MNRAS 442(2), 1545–1559. `arXiv:astro-ph.CO/1307.0943`.

Chabanier, S., F. Bournaud, Y. Dubois, S. Codis, D. Chapon, et al. (2020). Formation of compact galaxies in the Extreme-Horizon simulation. arXiv e-prints, arXiv:2007.04624. `arXiv:astro-ph.GA/2007.04624`.

Chabanier, S., F. Bournaud, Y. Dubois, N. Palanque-Delabrouille, C. Yèche, et al. (2020). The impact of AGN feedback on the 1D power spectra from the Lyα forest using the Horizon-AGN suite of simulations. MNRAS. arXiv:astro-ph.GA/2002.02822.

Chabanier, S., M. Millea, and N. Palanque-Delabrouille (2019). Matter power spectrum: from Ly α forest to CMB scales. MNRAS 489(2), 2247–2253. arXiv:astro-ph.CO/1905.08103.

Chabanier, S., N. Palanque-Delabrouille, C. Yèche, J.-M. Le Goff, E. Armengaud, et al. (2019). The one-dimensional power spectrum from the SDSS DR14 Lyα forests. J. Cosmology Astropart. Phys. 2019(7), 017. arXiv:astro-ph.CO/1812.03554.

Chartrand, R. (2005). Numerical differentiation of noisy, nonsmooth data. ISRN Applied Mathematics 2011, Article ID 164564.

Chisari, N. E., M. L. A. Richardson, J. Devriendt, Y. Dubois, A. Schneider, et al. (2018). The impact of baryons on the matter power spectrum from the Horizon-AGN cosmological hydrodynamical simulation. MNRAS 480(3), 3962–3977. arXiv:astro-ph.CO/1801.08559.

Clowe, D., M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, et al. (2006). A Direct Empirical Proof of the Existence of Dark Matter. Astrophysical Journal Letters 648(2), L109–L113. arXiv:astro-ph/astro-ph/0608407.

Colella, P. and P. R. Woodward (1984). The piecewise parabolic method (ppm) for gas-dynamical simulations. Journal of Computational Physics 54(1), 174 – 201.

Collaboration, D., A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, et al. (2016). The desi experiment part i: Science,targeting, and survey design.

Collaboration, P., N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, et al. (2018). Planck 2018 results. VIII. Gravitational lensing. arXiv:1807.06210 [astro-ph]. arXiv:astro-ph/1807.06210.

Collaboration, P., Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, et al. (2018). Planck 2018 results. I. Overview and the cosmological legacy of Planck. arXiv:1807.06205 [astro-ph]. arXiv:astro-ph/1807.06205.

Colless, M., G. Dalton, S. Maddox, W. Sutherland, P. Norberg, et al. (2001). The 2dF Galaxy Redshift Survey: spectra and redshifts. MNRAS 328, 1039–1063. astro-ph/0106498.

Colombi, S., S. Dodelson, and L. M. Widrow (1996). Large-Scale Structure Tests of Warm Dark Matter. Astrophysical Journal 458, 1. arXiv:astro-ph/astro-ph/9505029.

Courant, R., K. Friedrichs, and H. Lewy (1967). On the Partial Difference Equations of Mathematical Physics. IBM Journal of Research and Development 11, 215–234.

Cowan, C. L., J., F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire (1956). Detection of the Free Neutrino: A Confirmation. Science 124(3212), 103–104.

Crocce, M., S. Pueblas, and R. Scoccimarro (2006). Transients from initial conditions in cosmological simulations. MNRAS 373(1), 369–381. arXiv:astro-ph/astro-ph/0606505.

Croft, R. A. C., D. H. Weinberg, M. Bolte, S. Burles, L. Hernquist, et al. (2002). Toward a Precise Measurement of Matter Clustering: Lyα Forest Data at Redshifts 2-4. Astrophysical Journal 581, 20–52. astro-ph/0012324.

Croft, R. A. C., D. H. Weinberg, N. Katz, and L. Hernquist (1998). Cosmology from the structure of the lya forest. In V. Mueller, S. Gottloeber, J. P. Muecket, and J. Wambsganss (Eds.), Large Scale Structure: Tracks and Traces, pp. 69–75.

Croft, R. A. C., D. H. Weinberg, N. Katz, and L. Hernquist (1998). Recovery of the Power Spectrum of Mass Fluctuations from Observations of the Ly-alpha Forest. The Astrophysical Journal 495(1), 44–62. arXiv:astro-ph/9708018.

Croft, R. A. C., D. H. Weinberg, M. Pettini, L. Hernquist, and N. Katz (1999). The Power Spectrum of Mass Fluctuations Measured from the Ly-alpha Forest at Redshift z = 2.5. The Astrophysical Journal 520(1), 1–23. arXiv:astro-ph/9809401.

Croton, D. J. (2006). Evolution in the black hole mass?bulge mass relation: a theoretical perspective. Monthly Notices of the Royal Astronomical Society 369(4), 1808–1812. http://oup.prod.sis.lan/mnras/article-pdf/369/4/1808/3815593/mnras0369-1808.pdf.

Daddi, E., F. Bournaud, F. Walter, H. Dannerbauer, C. L. Carilli, et al. (2010). Very High Gas Fractions and Extended Gas Reservoirs in z = 1.5 Disk Galaxies. Astrophysical Journal 713(1), 686–707. arXiv:astro-ph.CO/0911.2776.

Daddi, E., A. Renzini, N. Pirzkal, A. Cimatti, S. Malhotra, et al. (2005). Passively Evolving Early-Type Galaxies at 1.4 <$\sim$z <$\sim$2.5 in the Hubble Ultra Deep Field. Astrophysical Journal 626(2), 680–697. arXiv:astro-ph/astro-ph/0503102.

Danovich, M., A. Dekel, O. Hahn, D. Ceverino, and J. Primack (2015). Four phases of angular-momentum buildup in high-z galaxies: from cosmic-web streams through an extended ring to disc and bulge. MNRAS 449(2), 2087–2111. arXiv:astro-ph.GA/1407.7129.

Davé, R., B. D. Oppenheimer, and K. Finlator (2011). Galaxy evolution in cosmological simulations with outflows - I. Stellar masses and star formation rates. MNRAS 415(1), 11–31. arXiv:astro-ph.CO/1103.3528.

Davis, M., G. Efstathiou, C. S. Frenk, and S. D. M. White (1985). The evolution of large-scale structure in a universe dominated by cold dark matter. Astrophysical Journal 292, 371–394.

Davis, R., D. S. Harmer, and K. C. Hoffman (1968). Search for Neutrinos from the Sun. Phys. Rev. Lett. 20(21), 1205–1209.

Dawson, K. S., J.-P. Kneib, W. J. Percival, S. Alam, F. D. Albareti, et al. (2016). The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data. Astronomical Journal 151, 44. arXiv:1508.04473.

Dawson, K. S., D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, et al. (2013). The Baryon Oscillation Spectroscopic Survey of SDSS-III. Astronomical Journal 145(1), 10. arXiv:astro-ph.CO/1208.0022.

de Bernardis, P., P. A. R. Ade, J. J. Bock, J. R. Bond, J. Borrill, et al. (2000). A flat Universe from high-resolution maps of the cosmic microwave background radiation. Nature 404(6781), 955–959. arXiv:astro-ph/astro-ph/0004404.

de Sainte Agathe, V., C. Balland, H. du Mas des Bourboux, N. G. Busca, M. Blomqvist, et al. (2019). Baryon acoustic oscillations at z = 2.34 from the correlations of lyα absorption in eboss dr14. Astronomy & Astrophysics 629, A85.

de Salas, P. F., D. V. Forero, C. A. Ternes, M. Tórtola, and J. W. F. Valle (2018). Status of neutrino oscillations 2018: 3σ hint for normal mass ordering and improved CP sensitivity. Physics Letters B 782, 633–640.

Decarli, R., R. Falomo, A. Treves, M. Labita, J. K. Kotilainen, et al. (2010). The quasar MBH-Mhost relation through cosmic time - II. Evidence for evolution from z = 3 to the present age. MNRAS 402(4), 2453–2461. arXiv:astro-ph.CO/0911.2988.

DeGraf, C., A. Dekel, J. Gabor, and F. Bournaud (2017). Black hole growth and AGN feedback under clumpy accretion. MNRAS 466(2), 1462–1476. arXiv:astro-ph.GA/1412.3819.

Dekel, A. and A. Burkert (2014). Wet disc contraction to galactic blue nuggets and quenching to red nuggets. MNRAS 438(2), 1870–1879. arXiv:astro-ph.CO/1310.1074.

Dekel, A. and J. Silk (1986). The Origin of Dwarf Galaxies, Cold Dark Matter, and Biased Galaxy Formation. Astrophysical Journal 303, 39.

Delubac, T., J. E. Bautista, N. G. Busca, J. Rich, D. Kirkby, et al. (2015). Baryon acoustic oscillations in the Lyα forest of BOSS DR11 quasars. Astronomy & Astrophysics 574, A59. arXiv:1404.1801.

Dessert, C., N. L. Rodd, and B. R. Safdi (2018). The dark matter interpretation of the 3.5-keV line is inconsistent with blank-sky observations. arXiv e-prints, arXiv:1812.06976. arXiv:astro-ph.CO/1812.06976.

Di Matteo, T., J. Colberg, V. Springel, L. Hernquist, and D. Sijacki (2008). Direct Cosmological Simulations of the Growth of Black Holes and Galaxies. Astrophysical Journal 676(1), 33–53. arXiv:astro-ph/0705.2269.

Dicke, R. H., P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson (1965). Cosmic Black-Body Radiation. Astrophysical Journal 142, 414–419.

Dinshaw, N., R. J. Weymann, C. D. Impey, C. B. Foltz, S. L. Morris, et al. (1997). Additional Observations and Analysis of the Lyman-α Absorption Lines toward the QSO Pair Q0107-025A,B. Astrophysical Journal 491(1), 45–68.

Dodelson, S. (2003). Modern cosmology.

Dodelson, S. and G. Efstathiou (2004). Modern Cosmology. Physics Today 57(7), 60–61.

Dodelson, S. and L. M. Widrow (1994). Sterile neutrinos as dark matter. Phys. Rev. Lett. 72(1), 17–20. arXiv:hep-ph/hep-ph/9303287.

Drewes, M. (2013). The Phenomenology of Right Handed Neutrinos. International Journal of Modern Physics E 22(8), 1330019–593. arXiv:hep-ph/1303.6912.

Drlica-Wagner, A. et al. (2018). Dark Energy Survey Year 1 Results: Photometric Data Set for Cosmology. Astrophys. J. Suppl. 235(2), 33. arXiv:astro-ph.CO/1708.01531.

Drui, F., A. Fikl, P. Kestener, S. Kokh, A. Larat, et al. (2016). Experimenting with the p4est library for amr simulations of two-phase flows. ArXiv abs/1709.07700.

du Mas des Bourboux, H., J.-M. Le Goff, M. Blomqvist, N. G. Busca, J. Guy, et al. (2017). Baryon acoustic oscillations from the complete sdss-iii lyα-quasar cross-correlation function at z = 2.4. Astronomy & Astrophysics 608, A130.

Dubois, Y., J. Devriendt, A. Slyz, and R. Teyssier (2012). Self-regulated growth of supermassive black holes by a dual jet-heating active galactic nucleus feedback mechanism: methods, tests and implications for cosmological simulations. MNRAS 420(3), 2662–2683. arXiv:astro-ph.CO/1108.0110.

Dubois, Y., S. Peirani, C. Pichon, J. Devriendt, R. Gavazzi, et al. (2016). The HORIZON-AGN simulation: morphological diversity of galaxies promoted by AGN feedback. MNRAS 463(4), 3948–3964. arXiv:astro-ph.GA/1606.03086.

Dubois, Y., M. Volonteri, and J. Silk (2014). Black hole evolution - III. Statistical properties of mass growth and spin evolution using large-scale hydrodynamical cosmological simulations. MNRAS 440(2), 1590–1606. arXiv:astro-ph.CO/1304.4583.

Dubois, Y., M. Volonteri, J. Silk, J. Devriendt, A. Slyz, et al. (2015). Black hole evolution - I. Supernova-regulated black hole growth. MNRAS 452(2), 1502–1518. arXiv:astro-ph.GA/1504.00018.

Dutton, A. A., F. C. van den Bosch, S. M. Faber, L. Simard, S. A. Kassin, et al. (2011). On the evolution of the velocity-mass-size relations of disc-dominated galaxies over the past 10 billion years. MNRAS 410(3), 1660–1676. arXiv:astro-ph.GA/1006.3558.

Dvali, G. (2004). Infrared Modification of Gravity. arXiv e-prints, hep–th/0402130. arXiv:hep-th/hep-th/0402130.

Dvorkin, C., K. Blum, and M. Kamionkowski (2014). Constraining dark matter-baryon scattering with linear cosmology. Phys. Rev. D 89(2), 023519. arXiv:astro-ph.CO/1311.2937.

Efstathiou, G., M. Davis, S. D. M. White, and C. S. Frenk (1985). Numerical techniques for large cosmological N-body simulations. Astrophysical Journal Supplement Series 57, 241–260.

Eisenstein, D. J., D. H. Weinberg, E. Agol, H. Aihara, C. A. Prieto, et al. (2011). Sdss-iii: Massive spectroscopic surveys of the distant universe, the milky way galaxy, and extra-solar planetary systems.

Eisenstein, D. J., I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, et al. (2005). Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. Astrophysical Journal 633(2), 560–574. arXiv:astro-ph/astro-ph/0501171.

Elbaz, D., R. Leiton, N. Nagar, K. Okumura, M. Franco, et al. (2018). Starbursts in and out of the star-formation main sequence. Astronomy & Astrophysics 616, A110. arXiv:astro-ph.GA/1711.10047.

Elbert, O. D., J. S. Bullock, S. Garrison-Kimmel, M. Rocha, J. Oñorbe, et al. (2015). Core formation in dwarf haloes with self-interacting dark matter: no fine-tuning necessary. MNRAS 453(1), 29–37. arXiv:astro-ph.GA/1412.1477.

Elmegreen, B. G., C. Struck, and D. A. Hunter (2014). Shrinking Galaxy Disks with Fountain-driven Accretion from the Halo. Astrophysical Journal 796(2), 110. arXiv:astro-ph.GA/1411.0332.

Fall, S. M. and G. Efstathiou (1980). Formation and rotation of disc galaxies with haloes. MNRAS 193, 189–206.

Faucher-Giguère, C.-A., A. Lidz, M. Zaldarriaga, and L. Hernquist (2009). A New Calculation of the Ionizing Background Spectrum and the Effects of He II Reionization. Astrophysical Journal 703(2), 1416–1443. arXiv:astro-ph.CO/0901.4554.

Feldman, G. J. and R. D. Cousins (1998). Unified approach to the classical statistical analysis of small signals. Physical Review D 57, 3873–3889. physics/9711021.

Feldmann, R., N. Y. Gnedin, and A. V. Kravtsov (2012). The X-factor in Galaxies. II. The Molecular-hydrogen-Star-formation Relation. Astrophysical Journal 758(2), 127. arXiv:astro-ph.CO/1204.3910.

Fensch, J. (2017). Star and stellar cluster formation in gas-dominated galaxies. Ph. D. thesis. ThÃšse de doctorat dirigÃ©e par Duc, Pierre-Alain Physique. Physique de l'Univers Sorbonne Paris CitÃ© 2017.

Ferrarese, L. and D. Merritt (2000). A Fundamental Relation between Supermassive Black Holes and Their Host Galaxies. Astrophysical Journal Letters 539(1), L9–L12. arXiv:astro-ph/astro-ph/0006053.

Ferreira, P. G. and M. Joyce (1997). Structure Formation with a Self-Tuning Scalar Field. Phys. Rev. Lett. 79(24), 4740–4743. arXiv:astro-ph/astro-ph/9707286.

Follin, B., L. Knox, M. Millea, and Z. Pan (2015). First Detection of the Acoustic Oscillation Phase Shift Expected from the Cosmic Neutrino Background. Phys. Rev. Lett. 115(9), 091301. arXiv:astro-ph.CO/1503.07863.

Font-Ribera, A., D. Kirkby, N. Busca, J. Miralda-Escudé, N. P. Ross, et al. (2014). Quasar-Lyman α forest cross-correlation from BOSS DR11: Baryon Acoustic Oscillations. J. Cosmology Astropart. Phys. 5, 027. arXiv:1311.1767.

Font-Ribera, A., P. McDonald, and J. Miralda-Escudé (2012). Generating mock data sets for large-scale Lyman-α forest correlation measurements. J. Cosmology Astropart. Phys. 1, 001. arXiv:1108.5606.

Font-Ribera, A., P. McDonald, and A. Slosar (2018). How to estimate the 3D power spectrum of the Lyman-α forest. J. Cosmology Astropart. Phys. 2018(1), 003. arXiv:astro-ph.CO/1710.11036.

Fontanot, F., G. De Lucia, P. Monaco, R. S. Somerville, and P. Santini (2009). The many manifestations of downsizing: hierarchical galaxy formation models confront observations. MNRAS 397(4), 1776–1790. arXiv:astro-ph.CO/0901.1130.

Förster Schreiber, N. M., H. Übler, R. L. Davies, R. Genzel, E. Wisnioski, et al. (2019). The KMOS^{3D} Survey: Demographics and Properties of Galactic Outflows at z = 0.6-2.7. Astrophysical Journal 875(1), 21. arXiv:astro-ph.GA/1807.04738.

Freedman, W. L., B. F. Madore, T. Hoyt, I. S. Jang, R. Beaton, et al. (2020). Calibration of the Tip of the Red Giant Branch. Astrophysical Journal 891(1), 57. arXiv:astro-ph.GA/2002.01550.

Fromang, S., P. Hennebelle, and R. Teyssier (2006). A high order Godunov scheme with constrained transport and adaptive mesh refinement for astrophysical magnetohydrodynamics. Astronomy & Astrophysics 457(2), 371–384. arXiv:astro-ph/astro-ph/0607230.

Fry, A. B., F. Governato, A. Pontzen, T. Quinn, M. Tremmel, et al. (2015). All about baryons: revisiting SIDM predictions at small halo masses. MNRAS 452(2), 1468–1479. arXiv:astro-ph.CO/1501.00497.

Fukuda, Y., T. Hayakawa, E. Ichihara, K. Inoue, K. Ishihara, et al. (1998). Evidence for Oscillation of Atmospheric Neutrinos. Phys. Rev. Lett. 81(8), 1562–1567. arXiv:hep-ex/hep-ex/9807003.

Gabor, J. M. and F. Bournaud (2014). Delayed star formation in high-redshift stream-fed galaxies. MNRAS 437(1), L56–L60. arXiv:astro-ph.CO/1310.1923.

Garny, M., T. Konstandin, L. Sagunski, and S. Tulin (2018). Lyman-α forest constraints on interacting dark sectors. J. Cosmology Astropart. Phys. 2018(9), 011. arXiv:astro-ph.CO/1805.12203.

Garrison, L. H., D. J. Eisenstein, D. Ferrer, M. V. Metchnik, and P. A. Pinto (2016). Improving initial conditions for cosmological N-body simulations. MNRAS 461(4), 4125–4145. arXiv:astro-ph.CO/1605.02333.

Garzilli, A., A. Magalich, T. Theuns, C. S. Frenk, C. Weniger, et al. (2019). The Lyman-α forest as a diagnostic of the nature of the dark matter. MNRAS 489(3), 3456–3471. arXiv:astro-ph.CO/1809.06585.

Garzilli, A., T. Theuns, and J. Schaye (2015). The broadening of Lyman-α forest absorption lines. MNRAS 450(2), 1465–1476. arXiv:astro-ph.CO/1502.05715.

Gebhardt, K., R. Bender, G. Bower, A. Dressler, S. M. Faber, et al. (2000). A Relationship between Nuclear Black Hole Mass and Galaxy Velocity Dispersion. Astrophysical Journal Letters 539(1), L13–L16. arXiv:astro-ph/astro-ph/0006289.

Gerritsen, J. P. E. and V. Icke (1997). Star formation in N-body simulations. I. The impact of the stellar ultraviolet radiation on star formation. Astronomy & Astrophysics 325, 972–986.

Gheller, C., P. Wang, F. Vazza, and R. Teyssier (2015). Numerical cosmology on the GPU with Enzo and Ramses. In Journal of Physics Conference Series, Volume 640 of Journal of Physics Conference Series, pp. 012058.

Gilbert, I. H. (1966). An Integral Equation for the Development of Irregularities in an Expanding Universe. Astrophysical Journal 144, 233.

Gnedin, N. Y. and A. J. S. Hamilton (2002). Matter power spectrum from the Lyman-alpha forest: Myth or reality? Mon. Not. Roy. Astron. Soc. 334, 107–116. arXiv:astro-ph/astro-ph/0111194.

Godunov, S. K. and V. S. Ryabenki (1964). Theory of difference schemes - an introduction.

Goerdt, T., D. Ceverino, A. Dekel, and R. Teyssier (2015). Distribution of streaming rates into high-redshift galaxies. MNRAS 454(1), 637–648. arXiv:astro-ph.CO/1505.01486.

Gontcho, S. G. A., J. Miralda-Escudé, and N. G. Busca (2014). On the effect of the ionising background on the Lyα forest autocorrelation function. ArXiv e-prints. arXiv:1404.7425.

Governato, F., B. Willman, L. Mayer, A. Brooks, G. Stinson, et al. (2007). Forming disc galaxies in ΛCDM simulations. MNRAS 374(4), 1479–1494. arXiv:astro-ph/astro-ph/0602351.

Guillet, T. and R. Teyssier (2011). A simple multigrid scheme for solving the Poisson equation with arbitrary domain boundaries. Journal of Computational Physics 230(12), 4756–4771. arXiv:physics.comp-ph/1104.1703.

Gunn, J. E., M. Carr, C. Rockosi, M. Sekiguchi, K. Berry, et al. (1998). The Sloan Digital Sky Survey Photometric Camera. Astronomical Journal 116(6), 3040–3081. arXiv:astro-ph/astro-ph/9809085.

Gunn, J. E. and B. A. Peterson (1965). On the Density of Neutral Hydrogen in Intergalactic Space. Astrophysical Journal 142, 1633–1636.

Gunn, J. E., W. A. Siegmund, E. J. Mannery, R. E. Owen, C. L. Hull, et al. (2006). The 2.5 m Telescope of the Sloan Digital Sky Survey. The Astrophysical Journal 131, 2332–2359. astro-ph/0602326.

Guo, Q., S. White, C. Li, and M. Boylan-Kolchin (2010). How do galaxies populate dark matter haloes? MNRAS 404(3), 1111–1120. arXiv:astro-ph.CO/0909.4305.

Haardt, F. and P. Madau (1996). Radiative Transfer in a Clumpy Universe. II. The Ultraviolet Extragalactic Background. Astrophysical Journal 461, 20. astro-ph/9509093.

Haardt, F. and P. Madau (2012). Radiative Transfer in a Clumpy Universe. IV. New Synthesis Models of the Cosmic UV/X-Ray Background. Astrophysical Journal 746(2), 125. arXiv:astro-ph.CO/1105.2039.

Habouzit, M., M. Volonteri, and Y. Dubois (2017). Blossoms from black hole seeds: properties and early growth regulated by supernova feedback. MNRAS 468(4), 3935–3948. arXiv:astro-ph.GA/1605.09394.

Hahn, O. and T. Abel (2011). Multi-scale initial conditions for cosmological simulations. MNRAS 415(3), 2101–2121. arXiv:astro-ph.CO/1103.6031.

Hanany, S., P. Ade, A. Balbi, J. Bock, J. Borrill, et al. (2000). MAXIMA-1: A Measurement of the Cosmic Microwave Background Anisotropy on Angular Scales of 10'-5 degree. Astrophysical Journal Letters 545(1), L5–L9. arXiv:astro-ph/astro-ph/0005123.

Häring, N. and H.-W. Rix (2004). On the Black Hole Mass-Bulge Mass Relation. Astrophysical Journal Letters 604(2), L89–L92. arXiv:astro-ph/astro-ph/0402376.

Harvey, D., R. Massey, T. Kitching, A. Taylor, and E. Tittley (2015). The nongravitational interactions of dark matter in colliding galaxy clusters. Science 347(6229), 1462–1465. arXiv:astro-ph.CO/1503.07675.

Hinshaw, G., D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, et al. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results. Astrophysical Journal Supplement Series 208, 19. arXiv:astro-ph.CO/1212.5226.

Hiss, H., M. Walther, J. Oñorbe, and J. F. Hennawi (2019). A Novel Statistical Method for Measuring the Temperature-Density Relation in the IGM Using the b - N $_{HI}$ Distribution of Absorbers in the Lyα Forest. Astrophysical Journal 876(1), 71. arXiv:astro-ph.CO/1903.11940.

Hjerting, F. (1938). Tables Facilitating the Calculation of Line Absorption Coefficients. Astrophysical Journal 88, 508.

Hlozek, R., J. Dunkley, G. Addison, J. W. Appel, J. R. Bond, et al. (2012). The Atacama Cosmology Telescope: A measurement of the primordial power spectrum. The Astrophysical Journal 749(1), 90. arXiv:1105.4887.

Hoekstra, H., H. K. C. Yee, and M. D. Gladders (2002). Constraints on Ω_m and σ_8 from Weak Lensing in Red-Sequence Cluster Survey Fields. Astrophysical Journal 577(2), 595–603. arXiv:astro-ph/astro-ph/0204295.

Hooper, D. and S. Profumo (2007). Dark matter and collider phenomenology of universal extra dimensions. Phys. Rep. 453(2-4), 29–115. arXiv:hep-ph/hep-ph/0701197.

Hopkins, P. F., E. Quataert, and N. Murray (2011). Self-regulated star formation in galaxies via momentum input from massive stars. MNRAS 417(2), 950–973. arXiv:astro-ph.CO/1101.4940.

Horiuchi, S., K. C. Y. Ng, J. M. Gaskins, M. Smith, and R. Preece (2015). Improved limits on sterile neutrino dark matter from full-sky observations by the Fermi-GBM. arXiv e-prints, arXiv:1502.03399. arXiv:astro-ph.HE/1502.03399.

Hou, Z., C. L. Reichardt, K. T. Story, B. Follin, R. Keisler, et al. (2014). Constraints on Cosmology from the Cosmic Microwave Background Power Spectrum of the 2500 deg^2 SPT-SZ Survey. Astrophysical Journal 782, 74. arXiv:astro-ph.CO/1212.6267.

Hu, E. M., T.-S. Kim, L. L. Cowie, A. Songaila, and M. Rauch (1995). The Distribution of Column Densities and B Values in the Lyman-Alpha Forest. Astronomical Journal 110, 1526. arXiv:astro-ph/astro-ph/9507047.

Hu, W., R. Barkana, and A. Gruzinov (2000). Fuzzy Cold Dark Matter: The Wave Properties of Ultralight Particles. Phys. Rev. Lett. 85(6), 1158–1161. arXiv:astro-ph/astro-ph/0003365.

Hubble, E. (1929). A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. Contributions from the Mount Wilson Observatory 3, 23–28.

Hui, L., N. Y. Gnedin, and Y. Zhang (1997). The Statistics of Density Peaks and the Column Density Distribution of the Lyα Forest. Astrophysical Journal 486(2), 599–622. arXiv:astro-ph/astro-ph/9608157.

Hui, L., A. Stebbins, and S. Burles (1999). A Geometrical Test of the Cosmological Energy Contents Using the Ly-alpha Forest. The Astrophysical Journal 511(1), L5–L8. arXiv:astro-ph/9807190.

Ikeuchi, S. (1986). The baryon clump within an extended dark matter region. Ap&SS 118(1-2), 509–514.

Iršič, V., M. Viel, M. G. Haehnelt, J. S. Bolton, and G. D. Becker (2017). First constraints on fuzzy dark matter from lyman-α forest data and hydrodynamical simulations. Phys. Rev. Lett. 119, 031302.

Iršič, V., M. Viel, T. A. M. Berg, V. D'Odorico, M. G. Haehnelt, et al. (2016). The Lyman-alpha forest power spectrum from the XQ-100 Legacy Survey. MNRAS.

Jenkins, A. (2010). Second-order Lagrangian perturbation theory initial conditions for resimulations. MNRAS 403(4), 1859–1872. arXiv:astro-ph.CO/0910.0258.

Jernigan, J. G. and D. H. Porter (1989). A Tree Code with Logarithmic Reduction of Force Terms, Hierarchical Regularization of All Variables, and Explicit Accuracy Controls. Astrophysical Journal Supplement Series 71, 871.

Joudaki, S. et al. (2019). KiDS+VIKING-450 and DES-Y1 combined: Cosmology with cosmic shear. arXiv:astro-ph.CO/1906.09262.

Joy, M., V. Sahni, and A. A. Starobinsky (2008). A New Universal Local Feature in the Inflationary Perturbation Spectrum. Phys. Rev. D77, 023514. arXiv:astro-ph/0711.1585.

Kaiser, N. (1987). Clustering in real space and in redshift space. MNRAS 227, 1–21.

Kajita, T. (1999). Atmospheric neutrino results from Super-Kamiokande and Kamiokande - Evidence for ν_μ oscillations. Nuclear Physics B Proceedings Supplements 77(1-3), 123–132. arXiv:hep-ex/hep-ex/9810001.

Kamkar, S. J., A. M. Wissink, V. Sankaran, and A. Jameson (2011). Feature-driven Cartesian adaptive mesh refinement for vortex-dominated flows. Journal of Computational Physics 230(16), 6271–6298.

Katz, N. (1992). Dissipational Galaxy Formation. II. Effects of Star Formation. Astrophysical Journal 391, 502.

Katz, N., D. Weinberg, L. Hernquist, and J. Miralda-Escudé (1996). Damped lyman-alpha and lyman-limit absorbers in the cold dark matter model. The Astrophysical Journal 457(2).

Katz, N., D. H. Weinberg, and L. Hernquist (1996). Cosmological simulations with TreeSPH. Astrophys. J. Suppl. 105, 19. arXiv:astro-ph/astro-ph/9509107.

Kaviraj, S., J. Devriendt, Y. Dubois, A. Slyz, C. Welker, et al. (2015). Galaxy merger histories and the role of merging in driving star formation atz > 1. Monthly Notices of the Royal Astronomical Society 452(3), 2845–2850.

Kennicutt, Jr., R. C. (1998). The Global Schmidt Law in Star-forming Galaxies. Astrophysical Journal 498, 541–552. astro-ph/9712213.

Khalatyan, A., A. Cattaneo, M. Schramm, S. Gottlöber, M. Steinmetz, et al. (2008). Is AGN feedback necessary to form red elliptical galaxies? MNRAS 387(1), 13–30. arXiv:astro-ph/0712.3289.

Khandai, N., T. Di Matteo, R. Croft, S. Wilkins, Y. Feng, et al. (2015). The massiveblack-ii simulation: the evolution of haloes and galaxies to z = 0. Monthly Notices of the Royal Astronomical Society 450(2), 1349–1374.

Khokhlov, A. (1998). Fully Threaded Tree Algorithms for Adaptive Refinement Fluid Dynamics Simulations. Journal of Computational Physics 143(2), 519–543. arXiv:astro-ph/astro-ph/9701194.

Kim, T. S., M. Viel, M. G. Haehnelt, R. F. Carswell, and S. Cristiani (2004). 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(2), 355–366. arXiv:astro-ph/0308103.

Kirkman, D., D. Tytler, D. Lubin, and J. Charlton (2007). Continuous statistics of the Lyα forest at 0 < z < 1.6: the mean flux, flux distribution and autocorrelation from HST FOS spectra. MNRAS 376(3), 1227–1237. arXiv:astro-ph/astro-ph/0612768.

Kobayashi, T. and F. Takahashi (2011). Running spectral index from inflation with modulations. JCAP 01, 026. arXiv:astro-ph.CO/1011.3988.

Kolb, E. W. and M. S. Turner (1990). The early universe.

Komatsu, E., K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, et al. (2011). Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. Astrophysical Journal Supplement Series 192, 18. arXiv:astro-ph.CO/1001.4538.

Kosowsky, A. and M. S. Turner (1995). CBR anisotropy and the running of the scalar spectral index. Phys. Rev. D 52, 1739. astro-ph/9504071.

Krajnović, D., M. Cappellari, P. T. de Zeeuw, and Y. Copin (2006). Kinemetry: a generalization of photometry to the higher moments of the line-of-sight velocity distribution. MNRAS 366(3), 787–802. arXiv:astro-ph/astro-ph/0512200.

Kraljic, K., F. Bournaud, and M. Martig (2012). The Two-phase Formation History of Spiral Galaxies Traced by the Cosmic Evolution of the Bar Fraction. Astrophysical Journal 757(1), 60. arXiv:astro-ph.GA/1207.0351.

Kriek, M., P. G. van Dokkum, M. Franx, G. D. Illingworth, and D. K. Magee (2009). The Hubble Sequence Beyond z = 2 for Massive Galaxies: Contrasting Large Star-forming and Compact Quiescent Galaxies. Astrophysical Journal Letters 705(1), L71–L75. arXiv:astro-ph.CO/0909.0260.

Krumholz, M. R. and J. C. Tan (2007). Slow Star Formation in Dense Gas: Evidence and Implications. Astrophysical Journal 654, 304–315. astro-ph/0606277.

Kuhlen, M., M. R. Krumholz, P. Madau, B. D. Smith, and J. Wise (2012). Dwarf Galaxy Formation with H_2-regulated Star Formation. Astrophysical Journal 749(1), 36. arXiv:astro-ph.CO/1105.2376.

Kusenko, A. (2006). Sterile Neutrinos, Dark Matter, and Pulsar Velocities in Models with a Higgs Singlet. Phys. Rev. Lett. 97(24), 241301. arXiv:hep-ph/hep-ph/0609081.

Lackner, C. N., R. Cen, J. P. Ostriker, and M. R. Joung (2012). Building galaxies by accretion and in situ star formation. MNRAS 425(1), 641–656. arXiv:astro-ph.CO/1206.0295.

Lahanas, A. B., N. E. Mavromatos, and D. V. Nanopoulos (2007). Smoothly evolving supercritical-string dark energy relaxes supersymmetric-dark-matter constraints. Physics Letters B 649(1), 83–90. arXiv:hep-ph/hep-ph/0612152.

Laigle, C. et al. (2016). The COSMOS2015 Catalog: Exploring the 1< z< 6 Universe with half a million galaxies. Astrophys. J. Suppl. 224(2), 24. arXiv:astro-ph.GA/1604.02350.

Laor, A. (2001). On the Linearity of the Black Hole-Bulge Mass Relation in Active and in Nearby Galaxies. Astrophysical Journal 553(2), 677–682. arXiv:astro-ph/astro-ph/0101405.

Larson, R. B. (1974). Effects of supernovae on the early evolution of galaxies. MNRAS 169, 229–246.

Lattanzi, M. and J. W. F. Valle (2007). Decaying Warm Dark Matter and Neutrino Masses. Phys. Rev. Lett. 99(12), 121301. arXiv:astro-ph/0705.2406.

Le Goff, J. M., C. Magneville, E. Rollinde, S. Peirani, P. Petitjean, et al. (2011). Simulations of BAO reconstruction with a quasar Ly-α survey. Astronomy & Astrophysics 534, A135. arXiv:1107.4233.

Lee, B. W. and S. Weinberg (1977). Cosmological lower bound on heavy-neutrino masses. Phys. Rev. Lett. 39(4), 165–168.

Lesgourgues, J. (2011). The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview. ArXiv e-prints. arXiv:astro-ph.IM/1104.2932.

Lesgourgues, J., G. Mangano, G. Miele, and S. Pastor (2013). Neutrino Cosmology. Cambridge University Press.

Lesgourgues, J. and S. Pastor (2006). Massive neutrinos and cosmology. Phys.Rept. 429, 307–379. arXiv:astro-ph/astro-ph/0603494.

Lewis, A., A. Challinor, and A. Lasenby (2000). Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models. The Astrophysical Journal 538(2), 473–476. arXiv:astro-ph/9911177.

L'Huillier, B., C. Park, and J. Kim (2014). Effects of the initial conditions on cosmological N-body simulations. New Astron. 30, 79–88. arXiv:astro-ph.CO/1401.6180.

Lilly, S. J. and C. M. Carollo (2016). Surface Density Effects in Quenching: Cause or Effect? Astrophysical Journal 833(1), 1. arXiv:astro-ph.GA/1604.06459.

Liske, J., J. K. Webb, G. M. Williger, A. Fernández-Soto, and R. F. Carswell (2000). Large-scale structure in the Lyman-α forest - II. Analysis of a group of 10 QSOs. MNRAS 311(4), 657–667. arXiv:astro-ph/astro-ph/9910407.

Lu, J.-L. and K.-L. Haung (1996). A statistical study of the Lyman-α forest. Chinese Astron. Astrophys. 20(2), 250–250.

Lukić, Z., C. W. Stark, P. Nugent, M. White, A. A. Meiksin, et al. (2015). The Lyman α forest in optically thin hydrodynamical simulations. MNRAS 446(4), 3697–3724. arXiv:astro-ph.CO/1406.6361.

Ma, C.-P. and E. Bertschinger (1994). Do Galactic Systems Form too Late in Cold + Hot Dark Matter Models? Astrophysical Journal Letters 434, L5. arXiv:astro-ph/astro-ph/9407085.

Madau, P., H. C. Ferguson, M. E. Dickinson, M. Giavalisco, C. C. Steidel, et al. (1996). High-redshift galaxies in the Hubble Deep Field: colour selection and star formation history to z$\sim$4. MNRAS 283(4), 1388–1404. arXiv:astro-ph/astro-ph/9607172.

Mandelker, N., D. Nagai, H. Aung, A. Dekel, Y. Birnboim, et al. (2020). Instability of supersonic cold streams feeding galaxies - IV. Survival of radiatively cooling streams. MNRAS 494(2), 2641–2663. arXiv:astro-ph.GA/1910.05344.

Marconi, A. and L. K. Hunt (2003). The Relation between Black Hole Mass, Bulge Mass, and Near-Infrared Luminosity. Astrophysical Journal Letters 589(1), L21–L24. arXiv:astro-ph/astro-ph/0304274.

Marsh, D. J. E. (2016). WarmAndFuzzy: the halo model beyond CDM. arXiv e-prints, arXiv:1605.05973. arXiv:astro-ph.CO/1605.05973.

Martig, M., F. Bournaud, D. J. Croton, A. Dekel, and R. Teyssier (2012). A Diversity of Progenitors and Histories for Isolated Spiral Galaxies. Astrophysical Journal 756(1), 26. arXiv:astro-ph.CO/1201.1079.

Martín-Navarro, I., J. N. Burchett, and M. Mezcua (2020). Black hole feedback and the evolution of massive early-type galaxies. MNRAS 491(1), 1311–1319. arXiv:astro-ph.GA/1910.12870.

Massara, E., F. Villaescusa-Navarro, M. Viel, and P. M. Sutter (2015). Voids in massive neutrino cosmologies. J. Cosmology Astropart. Phys. 2015(11), 018. arXiv:astro-ph.CO/1506.03088.

Matthews, T. A. (1963). Characteristics of Identified Extragalactic Radio Sources. Astronomical Journal 68, 77.

McAllister, L., E. Silverstein, and A. Westphal (2010). Gravity waves and linear inflation from axion monodromy. Phys.Rev.D 82, 046003. arXiv:hep-th/0808.0706.

McCarthy, I. G., J. Schaye, S. Bird, and A. M. C. Le Brun (2017). The BAHAMAS project: calibrated hydrodynamical simulations for large-scale structure cosmology. MNRAS 465(3), 2936–2965. arXiv:astro-ph.CO/1603.02702.

McCarthy, I. G., J. Schaye, T. J. Ponman, R. G. Bower, C. M. Booth, et al. (2010). The case for AGN feedback in galaxy groups. MNRAS 406(2), 822–839. `arXiv:astro-ph.CO/0911.2641`.

McDonald, P. (2003). Toward a Measurement of the Cosmological Geometry at z ˜ 2: Predicting Lyα Forest Correlation in Three Dimensions and the Potential of Future Data Sets. Astrophysical Journal 585, 34–51. `astro-ph/0108064`.

McDonald, P., J. Miralda-Escude, M. Rauch, W. L. W. Sargent, T. A. Barlow, et al. (2000). The Observed Probability Distribution Function, Power Spectrum, and Correlation Function of the Transmitted Flux in the Ly-alpha Forest. The Astrophysical Journal 543(1), 1–23. `arXiv:astro-ph/9911196`.

McDonald, P., U. Seljak, S. Burles, D. J. Schlegel, D. H. Weinberg, et al. (2006). The Lyα Forest Power Spectrum from the Sloan Digital Sky Survey. Astrophysical Journal Supplement Series 163(1), 80–109. `arXiv:astro-ph/astro-ph/0405013`.

McLure, R. J. and J. S. Dunlop (2002). On the black hole-bulge mass relation in active and inactive galaxies. MNRAS 331(3), 795–804. `arXiv:astro-ph/astro-ph/0108417`.

Meiksin, A. A. (2009). The physics of the intergalactic medium. Reviews of Modern Physics 81(4), 1405–1469.

Merle, A. (2013). keV Neutrino Model Building. International Journal of Modern Physics D 22(10), 1330020. `arXiv:hep-ph/1302.2625`.

Merle, A. and A. Schneider (2015). Production of Sterile Neutrino dark matter and the 3.5 keV line. Physics Letters B 749, 283–288. `arXiv:hep-ph/1409.6311`.

Merloni, A., A. Bongiorno, M. Bolzonella, M. Brusa, F. Civano, et al. (2010). On the Cosmic Evolution of the Scaling Relations Between Black Holes and Their Host Galaxies: Broad-Line Active Galactic Nuclei in the zCOSMOS Survey. Astrophysical Journal 708(1), 137–157. `arXiv:astro-ph.CO/0910.4970`.

Mocz, P. and S. Succi (2015). Numerical solution of the nonlinear Schrödinger equation using smoothed-particle hydrodynamics. Phys. Rev. E 91(5), 053304. `arXiv:physics.comp-ph/1503.03869`.

Monaghan, J. J. (1985). Extrapolating B. Splines for Interpolation. Journal of Computational Physics 60(2), 253–262.

Monaghan, J. J. and J. C. Lattanzio (1985). A refined particle method for astrophysical problems. Astronomy & Astrophysics 149, 135–143.

Mortlock, D. J., S. J. Warren, B. P. Venemans, M. Patel, P. C. Hewett, et al. (2011). A luminous quasar at a redshift of $z = 7.085$. Nature 474(7353), 616–619. `arXiv:astro-ph.CO/1106.6088`.

Moster, B. P., T. Naab, and S. D. M. White (2013). Galactic star formation and accretion histories from matching galaxies to dark matter haloes. MNRAS 428(4), 3121–3138. `arXiv:astro-ph.CO/1205.5807`.

Mould, J. R., J. P. Huchra, W. L. Freedman, J. Kennicutt, Robert C., L. Ferrarese, et al. (2000). Erratum: The Hubble Space Telescope Key Project on the Extragalactic Distance Scale. XXVIII. Combining the Constraints on the Hubble Constant. Astrophysical Journal 545(1), 547–547.

Mummery, B. O., I. G. McCarthy, S. Bird, and J. Schaye (2017). The separate and combined effects of baryon physics and neutrino free streaming on large-scale structure. MNRAS 471(1), 227–242. arXiv:astro-ph.CO/1702.02064.

Muratov, A. L., D. Kereš, C.-A. Faucher-Giguère, P. F. Hopkins, E. Quataert, et al. (2015). Gusty, gaseous flows of FIRE: galactic winds in cosmological simulations with explicit stellar feedback. MNRAS 454(3), 2691–2713. arXiv:astro-ph.GA/1501.03155.

Myers, A. D., N. Palanque-Delabrouille, A. Prakash, I. Pâris, C. Yeche, et al. (2015). The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Quasar Target Selection. Astrophysical Journal Supplement Series 221, 27. arXiv:1508.04472.

Narayanan, V. K., D. N. Spergel, R. Davé, and C.-P. Ma (2000). Constraints on the Mass of Warm Dark Matter Particles and the Shape of the Linear Power Spectrum from the Lyα Forest. Astrophysical Journal Letters 543(2), L103–L106. arXiv:astro-ph/astro-ph/0005095.

Nelson, D., A. Pillepich, V. Springel, R. Pakmor, R. Weinberger, et al. (2019). First results from the TNG50 simulation: galactic outflows driven by supernovae and black hole feedback. MNRAS 490(3), 3234–3261. arXiv:astro-ph.GA/1902.05554.

Neyman, J. (1937). Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. Trans. Royal Soc. London, Series A 236(3), 333.

Nilles, H. P. (1984). Supersymmetry, supergravity and particle physics. Phys. Rep. 110(1-2), 1–162.

Nori, M. and M. Baldi (2018). AX-GADGET: a new code for cosmological simulations of Fuzzy Dark Matter and Axion models. MNRAS 478(3), 3935–3951. arXiv:astro-ph.CO/1801.08144.

Nori, M., R. Murgia, V. Iršič, M. Baldi, and M. Viel (2019). Lyman α forest and non-linear structure characterization in Fuzzy Dark Matter cosmologies. MNRAS 482(3), 3227–3243. arXiv:astro-ph.CO/1809.09619.

Noterdaeme, P., P. Petitjean, W. C. Carithers, I. Pâris, A. Font-Ribera, et al. (2012). Column density distribution and cosmological mass density of neutral gas: Sloan Digital Sky Survey-III Data Release 9. Astronomy & Astrophysics 547, L1. arXiv:1210.1213.

Nusser, A. and M. Haehnelt (1999). A first step towards a direct inversion of the Lyman forest in QSO spectra. Monthly Notices of the Royal Astronomical Society 303(1), 179–187. arXiv:astro-ph/9806109.

Oñorbe, J., J. F. Hennawi, and Z. Lukić (2017). Self-consistent Modeling of Reionization in Cosmological Hydrodynamical Simulations. Astrophysical Journal 837(2), 106. arXiv:astro-ph.CO/1607.04218.

Ocvirk, P., C. Pichon, and R. Teyssier (2008). Bimodal gas accretion in the Horizon-MareNostrum galaxy formation simulation. MNRAS 390(4), 1326–1338. arXiv:astro-ph/0803.4506.

Oort, J. H. (1932). Note on the distribution of luminosities of K and M giants. Bull. Astron. Inst. Netherlands 6, 289.

Oppenheimer, B. D., R. Davé, D. Kereš, M. Fardal, N. Katz, et al. (2010). Feedback and recycled wind accretion: assembling the z = 0 galaxy mass function. MNRAS 406(4), 2325–2338. arXiv:astro-ph.CO/0912.0519.

Osato, K., T. Nishimichi, F. Bernardeau, and A. Taruya (2019). Perturbation theory challenge for cosmological parameters estimation: Matter power spectrum in real space. Phys. Rev. D 99(6), 063530. arXiv:astro-ph.CO/1810.10104.

Oser, L., J. P. Ostriker, T. Naab, P. H. Johansson, and A. Burkert (2010). The Two Phases of Galaxy Formation. Astrophysical Journal 725(2), 2312–2323. arXiv:astro-ph.CO/1010.1381.

O'Shea, B. W., K. Nagamine, V. Springel, L. Hernquist, and M. L. Norman (2005). Comparing AMR and SPH Cosmological Simulations. I. Dark Matter and Adiabatic Simulations. Astrophysical Journal Supplement Series 160(1), 1–27. arXiv:astro-ph/astro-ph/0312651.

Pagels, H. and J. R. Primack (1982). Supersymmetry, cosmology, and new physics at teraelectronvolt energies. Phys. Rev. Lett. 48, 223–226.

Palanque-Delabrouille, N., C. Magneville, C. Yèche, S. Eftekharzadeh, A. D. Myers, et al. (2013). Luminosity function from dedicated SDSS-III and MMT data of quasars in 0.7 < z < 4.0 selected with a new approach. A. & A. 551, A29. arXiv:1209.3968.

Palanque-Delabrouille, N., C. Magneville, C. Yèche, I. Pâris, P. Petitjean, et al. (2016). The extended Baryon Oscillation Spectroscopic Survey: Variability selection and quasar luminosity function. Astronomy & Astrophysics 587, A41. arXiv:1509.05607.

Palanque-Delabrouille, N., C. Yèche, J. Baur, C. Magneville, G. Rossi, et al. (2015). Neutrino masses and cosmology with lyman-alpha forest power spectrum. Journal of Cosmology and Astroparticle Physics 2015(11), 011–011.

Palanque-Delabrouille, N., C. Yèche, A. Borde, J.-M. Le Goff, G. Rossi, et al. (2013). The one-dimensional Ly-alpha forest power spectrum from BOSS. A. & A. 559(1), A85. arXiv:1306.5896.

Palanque-Delabrouille, N., C. Yèche, J. Lesgourgues, G. Rossi, A. Borde, et al. (2015). Constraint on neutrino masses from sdss-iii/boss lyα forest and other cosmological probes. Journal of Cosmology and Astroparticle Physics 2015(02), 045–045.

Palanque-Delabrouille, N., C. Yèche, J. Lesgourgues, G. Rossi, A. Borde, et al. (2015). Constraint on neutrino masses from SDSS-III/BOSS Lyα forest and other cosmological probes. J. Cosmology Astropart. Phys. 2, 45. arXiv:1410.7244.

Palanque-Delabrouille, N., C. Yèche, N. Schöneberg, J. Lesgourgues, M. Walther, et al. (2020). Hints, neutrino bounds, and WDM constraints from SDSS DR14 Lyman-α and Planck full-survey data. J. Cosmology Astropart. Phys. 2020(4), 038. arXiv:astro-ph.CO/1911.09073.

Pâris, I., P. Petitjean, E. Aubourg, S. Bailey, N. P. Ross, et al. (2012). The Sloan Digital Sky Survey quasar catalog: ninth data release. Astronomy & Astrophysics 548(October), A66. arXiv:1210.5166.

Pâris, I., P. Petitjean, É. Aubourg, A. D. Myers, A. Streblyanska, et al. (2018). The Sloan Digital Sky Survey Quasar Catalog: Fourteenth data release. Astronomy & Astrophysics 613, A51. arXiv:1712.05029.

Paris, I., P. Petitjean, E. Rollinde, E. Aubourg, N. Busca, et al. (2011). VizieR Online Data Catalog: Principal Component Analysis of QSO UV spectra (Paris+, 2011). VizieR Online Data Catalog 353.

Parks, D., J. X. Prochaska, S. Dong, and Z. Cai (2018). Deep learning of quasar spectra to discover and characterize damped Lyα systems. MNRAS 476, 1151–1168. arXiv:1709.04962.

Pedersen, C., A. Font-Ribera, T. D. Kitching, P. McDonald, S. Bird, et al. (2019). Massive neutrinos and degeneracies in Lyman-alpha forest simulations. arXiv e-prints, arXiv:1911.09596. arXiv:astro-ph.CO/1911.09596.

Penzias, A. A. and R. W. Wilson (1965a). A Measurement of Excess Antenna Temperature at 4080 Mc/s. Astrophysical Journal 142, 419–421.

Penzias, A. A. and R. W. Wilson (1965b). Measurement of the Flux Density of CAS a at 4080 Mc/s. Astrophysical Journal 142, 1149.

Pérez-Ràfols, I., A. Font-Ribera, J. Miralda-Escudé, M. Blomqvist, S. Bird, et al. (2017). 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(3), 3019–3038.

Perlmutter, S. (1999). Cosmological Parameters from Type Ia Supernovae at High Redshift. HST Proposal.

Péroux, C., R. G. McMahon, L. J. Storrie-Lombardi, and M. J. Irwin (2003). The evolution of Ω_{HI} and the epoch of formation of damped Lyman α absorbers. MNRAS 346(4), 1103–1115. arXiv:astro-ph/astro-ph/0107045.

Peskin, M. E. (2015). Supersymmetric dark matter in the harsh light of the Large Hadron Collider. Proceedings of the National Academy of Science 112(40), 12256–12263.

Peter, A. H. G., M. Rocha, J. S. Bullock, and M. Kaplinghat (2013). Cosmological simulations with self-interacting dark matter - II. Halo shapes versus observations. MNRAS 430(1), 105–120. arXiv:astro-ph.CO/1208.3026.

Petraki, K. and A. Kusenko (2008). Dark-matter sterile neutrinos in models with a gauge singlet in the Higgs sector. Phys. Rev. D 77(6), 065014. arXiv:hep-ph/0711.4646.

Pichon, C., D. Pogosyan, T. Kimm, A. Slyz, J. Devriendt, et al. (2011). Rigging dark haloes: why is hierarchical galaxy formation consistent with the inside-out build-up of thin discs? MNRAS 418(4), 2493–2507. arXiv:astro-ph.CO/1105.0210.

Pieri, M. M., S. Bonoli, J. Chaves-Montero, I. Paris, M. Fumagalli, et al. (2016). Weave-qso: A massive intergalactic medium survey for the william herschel telescope.

Pillepich, A., D. Nelson, V. Springel, R. Pakmor, P. Torrey, et al. (2019). First results from the TNG50 simulation: the evolution of stellar and gaseous discs across cosmic time. MNRAS 490(3), 3196–3233. arXiv:astro-ph.GA/1902.05553.

Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, et al. (2015). Planck 2015 results. XIII. Cosmological parameters. ArXiv:1502.01589. arXiv:1502.01589.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, et al. (2018a). Planck 2018 results. VI. Cosmological parameters. arXiv e-prints, arXiv:1807.06209. arXiv:astro-ph.CO/1807.06209.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, et al. (2018b). Planck 2018 results. VI. Cosmological parameters. ArXiv e-prints. arXiv:1807.06209.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, et al. (2018c). Planck 2018 results. VI. Cosmological parameters. arXiv:1807.06209 [astro-ph]. arXiv:1807.06209.

Pontecorvo, B. (1968). Neutrino Experiments and the Problem of Conservation of Leptonic Charge. Soviet Journal of Experimental and Theoretical Physics 26, 984.

Potter, D., J. Stadel, and R. Teyssier (2017). PKDGRAV3: beyond trillion particle cosmological simulations for the next era of galaxy surveys. Computational Astrophysics and Cosmology 4(1), 2. arXiv:astro-ph.IM/1609.08621.

Press, W. H. and P. Schechter (1974). Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. Astrophysical Journal 187, 425–438.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992). Numerical recipes in C. The art of scientific computing.

Prunet, S., C. Pichon, D. Aubert, D. Pogosyan, R. Teyssier, et al. (2008). Initial Conditions For Large Cosmological Simulations. Astrophysical Journal Supplement Series 178(2), 179–188. arXiv:astro-ph/0804.3536.

Puchwein, E., D. Sijacki, and V. Springel (2008). Simulations of AGN Feedback in Galaxy Clusters and Groups: Impact on Gas Fractions and the L_X-T Scaling Relation. Astrophysical Journal Letters 687(2), L53. arXiv:astro-ph/0808.0494.

Ratra, B. and P. J. E. Peebles (1988). Cosmological consequences of a rolling homogeneous scalar field. Phys. Rev. D 37(12), 3406–3427.

Rees, M. J. (1986). Lyman absorption lines in quasar spectra - Evidence for gravitationally-confined gas in dark minihaloes. MNRAS 218, 25P–30P.

Reid, B. A., W. J. Percival, D. J. Eisenstein, L. Verde, D. N. Spergel, et al. (2010). Cosmological constraints from the clustering of the Sloan Digital Sky Survey DR7 luminous red galaxies. Monthly Notices of the Royal Astronomical Society 404, 60–85.

Reisenegger, A. and J. Miralda-Escude (1995). The Gunn-Peterson Effect from Underdense Regions in a Photoionized Intergalactic Medium. Astrophysical Journal 449, 476. arXiv:astro-ph/astro-ph/9502063.

Ricarte, A., M. Tremmel, P. Natarajan, and T. Quinn (2019). Tracing black hole and galaxy co-evolution in the Romulus simulations. Monthly Notices of the Royal Astronomical Society 489(1), 802–819. https://academic.oup.com/mnras/article-pdf/489/1/802/29218683/stz2161.pdf.

Rich, J. (2001). Fundamentals of Cosmology.

Riess, A. G., S. Casertano, W. Yuan, L. M. Macri, and D. Scolnic (2019). Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM. Astrophys. J. 876(1), 85. arXiv:astro-ph.CO/1903.07603.

Riess, A. G., A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, et al. (1998). Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. Astronomical Journal 116(3), 1009–1038. arXiv:astro-ph/astro-ph/9805201.

Riess, A. G., L. M. Macri, S. L. Hoffmann, D. Scolnic, S. Casertano, et al. (2016). A 2.4% Determination of the Local Value of the Hubble Constant. Astrophysical Journal 826(1), 56. arXiv:astro-ph.CO/1604.01424.

Ringwald, A. (2012). Exploring the role of axions and other WISPs in the dark universe. Physics of the Dark Universe 1(1-2), 116–135. arXiv:hep-ph/1210.5081.

Ringwald, A. and Y. Y. Y. Wong (2004). Gravitational clustering of relic neutrinos and implications for their detection. J. Cosmology Astropart. Phys. 2004(12), 005. arXiv:hep-ph/hep-ph/0408241.

Robertson, B. E., A. V. Kravtsov, N. Y. Gnedin, T. Abel, and D. H. Rudd (2010). Computational Eulerian hydrodynamics and Galilean invariance. MNRAS 401(4), 2463–2476. arXiv:astro-ph.CO/0909.0513.

Rocha, M., A. H. G. Peter, J. S. Bullock, M. Kaplinghat, S. Garrison-Kimmel, et al. (2013). Cosmological simulations with self-interacting dark matter - I. Constant-density cores and substructure. MNRAS 430(1), 81–104. arXiv:astro-ph.CO/1208.3025.

Rodriguez-Gomez, V., A. Pillepich, L. V. Sales, S. Genel, M. Vogelsberger, et al. (2016). The stellar mass assembly of galaxies in the Illustris simulation: growth by mergers and the spatial distribution of accreted stars. MNRAS 458(3), 2371–2390. arXiv:astro-ph.GA/1511.08804.

Rogers, K. K., S. Bird, H. V. Peiris, A. Pontzen, A. Font-Ribera, et al. (2018). Simulating the effect of high column density absorbers on the one-dimensional Lyman α forest flux power spectrum. MNRAS 474, 3032–3042. arXiv:1706.08532.

Rorai, A., R. F. Carswell, M. G. Haehnelt, G. D. Becker, J. S. Bolton, et al. (2018). A new measurement of the intergalactic temperature at z $\sim$ 2.55-2.95. MNRAS 474(3), 2871–2883. arXiv:astro-ph.CO/1711.00930.

Rosdahl, J., J. Blaizot, D. Aubert, T. Stranex, and R. Teyssier (2013). RAMSES-RT: radiation hydrodynamics in the cosmological context. MNRAS 436(3), 2188–2231. arXiv:astro-ph.CO/1304.7126.

Ross, A. J. et al. (2014). The Clustering of Galaxies in the SDSS-III DR10 Baryon Oscillation Spectroscopic Survey: No Detectable Colour Dependence of Distance Scale or Growth Rate Measurements. Mon. Not. Roy. Astron. Soc. 437(2), 1109–1126. arXiv:astro-ph.CO/1310.1106.

Ross, N. P., A. D. Myers, E. S. Sheldon, C. Yèche, M. A. Strauss, et al. (2012). The SDSS-III Baryon Oscillation Spectroscopic Survey: Quasar Target Selection for Data Release Nine. The Astrophysical Journal Suppl. 199, 3. arXiv:astro-ph.CO/1105.0606.

Rossi, G., N. Palanque-Delabrouille, A. Borde, M. Viel, C. Yèche, et al. (2014). Suite of hydrodynamical simulations for the Lyman-α forest with massive neutrinos. Astronomy & Astrophysics 567, A79. arXiv:astro-ph.CO/1401.6464.

Rubin, V. C. and J. Ford, W. Kent (1970). Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. Astrophysical Journal 159, 379.

Ryan-Weber, E. V. (2006). Cross-correlation of Lyman-alpha absorbers with gas-rich galaxies. Monthly Notices of the Royal Astronomical Society 367(3), 1251–1260. https://academic.oup.com/mnras/article-pdf/367/3/1251/3522044/367-3-1251.pdf.

Saad, Y. (2003). Iterative Methods for Sparse Linear Systems (2nd ed.). USA: Society for Industrial and Applied Mathematics.

Santini, P., A. Fontana, A. Grazian, S. Salimbeni, F. Fontanot, et al. (2012). The evolving slope of the stellar mass function at 0.6 $\leq$ z < 4.5 from deep WFC3 data. Astronomy & Astrophysics 538, A33. arXiv:astro-ph.CO/1111.5728.

Sargent, W. L. W., P. J. Young, A. Boksenberg, and D. Tytler (1980). The distribution of Lyman-alpha absorption lines in the spectra of six QSOs: evidence for an intergalactic origin. Astrophysical Journal Supplement Series 42, 41–81.

Saslaw, W. C. (1985). Gravitational physics of stellar and galactic systems C. Saslaw.

Scannapieco, C., P. B. Tissera, S. D. M. White, and V. Springel (2008). Effects of supernova feedback on the formation of galaxy discs. MNRAS 389(3), 1137–1149. arXiv:astro-ph/0804.3795.

Scannapieco, C., M. Wadepuhl, O. H. Parry, J. F. Navarro, A. Jenkins, et al. (2012). The Aquila comparison project: the effects of feedback and numerical methods on simulations of galaxy formation. MNRAS 423(2), 1726–1749. arXiv:astro-ph.GA/1112.0315.

Scannapieco, E., C. Pichon, B. Aracil, P. Petitjean, R. J. Thacker, et al. (2006). MNRAS 365(2), 615–637. arXiv:astro-ph/astro-ph/0503001.

Schaye, J., C. Dalla Vecchia, C. M. Booth, R. P. C. Wiersma, T. Theuns, et al. (2010). The physics driving the cosmic star formation history. MNRAS 402(3), 1536–1560. arXiv:astro-ph.CO/0909.5196.

Schaye, J., T. Theuns, A. Leonard, and G. Efstathiou (1999). Measuring the equation of state of the intergalactic medium. MNRAS 310(1), 57–70. arXiv:astro-ph/astro-ph/9906271.

Scheuer, P. A. G. (1965). A Sensitive Test for the Presence of Atomic Hydrogen in Intergalactic Space. Nature 207(5000), 963.

Schneider, A., R. Teyssier, D. Potter, J. Stadel, J. Onions, et al. (2016). Matter power spectrum and the challenge of percent accuracy. J. Cosmology Astropart. Phys. 2016(4), 047. arXiv:astro-ph.CO/1503.05920.

Schreiber, C., M. Pannella, R. Leiton, D. Elbaz, T. Wang, et al. (2017). The ALMA Redshift 4 Survey (AR4S). I. The massive end of the z = 4 main sequence of galaxies. Astronomy & Astrophysics 599, A134. arXiv:astro-ph.GA/1606.06252.

Schroetter, I., N. F. Bouché, J. Zabl, T. Contini, M. Wendt, et al. (2019). MusE GAs FLOw and Wind (MEGAFLOW) - III. Galactic wind properties using background quasars. MNRAS 490(3), 4368–4381. arXiv:astro-ph.GA/1907.09967.

Schumann, M. (2019). Direct detection of WIMP dark matter: concepts and status. Journal of Physics G Nuclear Physics 46(10), 103003. arXiv:astro-ph.CO/1903.03026.

Seljak, U., A. Makarov, P. McDonald, and H. Trac (2006). Can Sterile Neutrinos Be the Dark Matter? Phys. Rev. Lett. 97(19), 191303. arXiv:astro-ph/astro-ph/0602430.

Seljak, U., P. McDonald, and A. Makarov (2003). Cosmological constraints from the cosmic microwave background and Lyman α forest revisited. MNRAS 342(4), L79–L84. arXiv:astro-ph/astro-ph/0302571.

Senatore, L. and M. Zaldarriaga (2017). The Effective Field Theory of Large-Scale Structure in the presence of Massive Neutrinos. arXiv e-prints, arXiv:1707.04698. arXiv:astro-ph.CO/1707.04698.

Shan, H., X. Liu, H. Hildebrandt, C. Pan, N. Martinet, et al. (2018). KiDS-450: cosmological constraints from weak lensing peak statistics - I. Inference from analytical prediction of high signal-to-noise ratio convergence peaks. MNRAS 474(1), 1116–1134. arXiv:astro-ph.CO/1709.07651.

Shaposhnikov, M. and I. Tkachev (2006). The νMSM, inflation, and dark matter. Physics Letters B 639(5), 414–417. arXiv:hep-ph/hep-ph/0604236.

Shi, X. and G. M. Fuller (1999). New Dark Matter Candidate: Nonthermal Sterile Neutrinos. Phys. Rev. Lett. 82(14), 2832–2835. arXiv:astro-ph/astro-ph/9810076.

Shklovskii, I. S. (1965). A Possible Secular Change in the Flux and Spectrum of the Radio Source 1934-63. Azh 42, 30.

Shoji, M. and E. Komatsu (2010). Erratum: Massive neutrinos in cosmology: Analytic solutions and fluid approximation [Phys. Rev. D 81, 123516 (2010)]. Phys. Rev. D 82(8), 089901. arXiv:astro-ph.CO/1003.0942.

Sievers, J. L., R. A. Hlozek, M. R. Nolta, V. Acquaviva, G. E. Addison, et al. (2013). The Atacama Cosmology Telescope: cosmological parameters from three seasons of data. J. Cosmology Astropart. Phys. 10, 60. arXiv:astro-ph.CO/1301.0824.

Sijacki, D., V. Springel, T. Di Matteo, and L. Hernquist (2007). A unified model for AGN feedback in cosmological simulations of structure formation. MNRAS 380(3), 877–900. arXiv:astro-ph/0705.2238.

Sijacki, D., M. Vogelsberger, S. Genel, V. Springel, P. Torrey, et al. (2015). The Illustris simulation: the evolving population of black holes across cosmic time. MNRAS 452(1), 575–596. arXiv:astro-ph.GA/1408.6842.

Silk, J. and M. J. Rees (1998). Quasars and galaxy formation. Astronomy & Astrophysics 331, L1–L4. arXiv:astro-ph/astro-ph/9801013.

Singh, S. and C.-P. Ma (2003). Neutrino clustering in cold dark matter halos: Implications for ultrahigh energy cosmic rays. Phys. Rev. D 67(2), 023506. arXiv:astro-ph/astro-ph/0208419.

Slosar, A., A. Font-Ribera, M. M. Pieri, J. Rich, J.-M. L. Goff, et al. (2011). The lyman-α forest in three dimensions: measurements of large scale flux correlations from boss 1st-year data. Journal of Cosmology and Astroparticle Physics 2011(09), 001–001.

Slosar, A., V. Iršič, D. Kirkby, S. Bailey, N. G. Busca, et al. (2013). Measurement of baryon acoustic oscillations in the lyman-α forest fluctuations in boss data release 9. Journal of Cosmology and Astroparticle Physics 2013(04), 026–026.

Smee, S. A., J. E. Gunn, A. Uomoto, N. Roe, D. Schlegel, et al. (2013). The Multi-object, Fiber-fed Spectrographs for the Sloan Digital Sky Survey and the Baryon Oscillation Spectroscopic Survey. The Astrophysical Journal 146, 32. arXiv:astro-ph.IM/1208.2233.

Smette, A., J. Surdej, P. A. Shaver, C. B. Foltz, F. H. Chaffee, et al. (1992). A Spectroscopic Study of UM 673 A and B: On the Size of Lyman-Alpha Clouds. Astrophysical Journal 389, 39.

Smoot, G. F., C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, et al. (1992). Structure in the COBE Differential Microwave Radiometer First-Year Maps. Astrophysical Journal Letters 396, L1.

Sommer-Larsen, J., M. Götz, and L. Portinari (2003). Galaxy Formation: Cold Dark Matter, Feedback, and the Hubble Sequence. Astrophysical Journal 596(1), 47–66. arXiv:astro-ph/astro-ph/0204366.

Sousbie, T. (2011). The persistent cosmic web and its filamentary structure - I. Theory and implementation. MNRAS 414, 350–383. arXiv:1009.4015.

Sousbie, T. (2013). DisPerSE: robust structure identification in 2D and 3D. ArXiv e-prints. arXiv:astro-ph.CO/1302.6221.

Spergel, D. N. and P. J. Steinhardt (2000). Observational Evidence for Self-Interacting Cold Dark Matter. Phys. Rev. Lett. 84(17), 3760–3763. arXiv:astro-ph/astro-ph/9909386.

Springel, V. (2005). The cosmological simulation code GADGET-2. Monthly Notices of the Royal Astronomical Society 364(4), 1105–1134. arXiv:astro-ph/0505010.

Springel, V. (2011). Moving-mesh hydrodynamics with the AREPO code. In J. Alves, B. G. Elmegreen, J. M. Girart, and V. Trimble (Eds.), Computational Star Formation, Volume 270 of IAU Symposium, pp. 203–206.

Springel, V., T. Di Matteo, and L. Hernquist (2005). Modelling feedback from stars and black holes in galaxy mergers. MNRAS 361(3), 776–794. `arXiv:astro-ph/astro-ph/0411108`.

Springel, V. and L. Hernquist (2003). Cosmological smoothed particle hydrodynamics simulations: a hybrid multiphase model for star formation. MNRAS 339(2), 289–311. `arXiv:astro-ph/astro-ph/0206393`.

Stinson, G. S., C. Brook, A. V. Macciò, J. Wadsley, T. R. Quinn, et al. (2013). Making Galaxies In a Cosmological Context: the need for early stellar feedback. MNRAS 428(1), 129–140. `arXiv:astro-ph.CO/1208.0002`.

Strafella, L. and D. Chapon (2020). Boosting I/O and visualization for exascale era using Hercule: test case on RAMSES. arXiv e-prints, arXiv:2006.02759. `arXiv:cs.DC/2006.02759`.

Sutherland, R. S. and M. A. Dopita (1993). Cooling functions for low-density astrophysical plasmas. Astrophysical Journal Supplement Series 88, 253–327.

Tabor, G. and J. Binney (1993). Elliptical galaxy cooling flows without mass drop-out. MNRAS 263, 323–334.

Tacchella, S., A. Dekel, C. M. Carollo, D. Ceverino, C. DeGraf, et al. (2016). Evolution of density profiles in high-z galaxies: compaction and quenching inside-out. MNRAS 458(1), 242–263. `arXiv:astro-ph.GA/1509.00017`.

Tacconi, L. J., R. Genzel, A. Saintonge, F. Combes, S. García-Burillo, et al. (2018). PHIBSS: Unified Scaling Relations of Gas Depletion Time and Molecular Gas Fractions. Astrophysical Journal 853(2), 179. `arXiv:astro-ph.GA/1702.01140`.

Tanabashi, M., K. Hagiwara, K. Hikasa, K. Nakamura, Y. Sumino, et al. (2018). Review of particle physics. Phys. Rev. D 98, 030001.

Tegmark, M. and M. Zaldarriaga (2002). Separating the early universe from the late universe: Cosmological parameter estimation beyond the black box. Phys. Rev. D 66(10), 103508. `arXiv:astro-ph/astro-ph/0207047`.

Tegmark, M. and M. Zaldarriaga (2009). The Fast Fourier Transform Telescope. Physical Review D 79(8), 083530. `arXiv:0805.4414`.

Tepper-García, T. (2006). Voigt profile fitting to quasar absorption lines: an analytic approximation to the Voigt-Hjerting function. MNRAS 369, 2025–2035. `astro-ph/0602124`.

Teyssier, R. (2002). Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES. Astronomy & Astrophysics 385, 337–364. `astro-ph/0111367`.

Teyssier, R., D. Chapon, and F. Bournaud (2010). The Driving Mechanism of Starbursts in Galaxy Mergers. Astrophysical Journal Letters 720(2), L149–L154. `arXiv:astro-ph.CO/1006.4757`.

Teyssier, R., B. Moore, D. Martizzi, Y. Dubois, and L. Mayer (2011). Mass distribution in galaxy clusters: the role of Active Galactic Nuclei feedback. MNRAS 414(1), 195–208. arXiv:astro-ph.CO/1003.4744.

Teyssier, R., A. Pontzen, Y. Dubois, and J. I. Read (2013). Cusp-core transformations in dwarf galaxies: observational predictions. MNRAS 429(4), 3068–3078. arXiv:astro-ph.CO/1206.4895.

Thacker, R. J. and H. M. P. Couchman (2000). Implementing Feedback in Simulations of Galaxy Formation: A Survey of Methods. Astrophysical Journal 545(2), 728–752. arXiv:astro-ph/astro-ph/0001276.

The Dark Energy Survey Collaboration (2005). The Dark Energy Survey. arXiv e-prints, astro–ph/0510346. arXiv:astro-ph/astro-ph/0510346.

Theuns, T., A. Leonard, G. Efstathiou, F. R. Pearce, and P. A. Thomas (1998). P**3M-SPH simulations of the lyman-alpha forest. Mon. Not. Roy. Astron. Soc. 301, 478–502. arXiv:astro-ph/astro-ph/9805119.

Tillson, H., J. Devriendt, A. Slyz, L. Miller, and C. Pichon (2015). Angular momentum transfer to a Milky Way disc at high redshift. MNRAS 449(4), 4363–4379. arXiv:astro-ph.CO/1211.3124.

Tomassetti, M., A. Dekel, N. Mandelker, D. Ceverino, S. Lapiner, et al. (2016). Evolution of galaxy shapes from prolate to oblate through compaction events. MNRAS 458(4), 4477–4497. arXiv:astro-ph.GA/1512.06268.

Tomczak, A. R., R. F. Quadri, K.-V. H. Tran, I. Labbé, C. M. S. Straatman, et al. (2014). Galaxy Stellar Mass Functions from ZFOURGE/CANDELS: An Excess of Low-mass Galaxies since z = 2 and the Rapid Buildup of Quiescent Galaxies. Astrophysical Journal 783(2), 85. arXiv:astro-ph.CO/1309.5972.

Toro, E. (2009). Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction.

Tremaine, S., K. Gebhardt, R. Bender, G. Bower, A. Dressler, et al. (2002). The Slope of the Black Hole Mass versus Velocity Dispersion Correlation. Astrophysical Journal 574(2), 740–753. arXiv:astro-ph/astro-ph/0203468.

Tremaine, S. and J. E. Gunn (1979). Dynamical role of light neutral leptons in cosmology. Phys. Rev. Lett. 42(6), 407–410.

Tremmel, M., M. Karcher, F. Governato, M. Volonteri, T. R. Quinn, et al. (2017). The Romulus cosmological simulations: a physical approach to the formation, dynamics and accretion models of SMBHs. MNRAS 470(1), 1121–1139. arXiv:astro-ph.GA/1607.02151.

Troxel, M. A., N. MacCrann, J. Zuntz, T. F. Eifler, E. Krause, et al. (2017). Dark Energy Survey Year 1 Results: Cosmological Constraints from Cosmic Shear. arXiv:1708.01538 [astro-ph]. arXiv:astro-ph/1708.01538.

Truelove, J. K., R. I. Klein, C. F. McKee, J. H. Holliman, II, L. H. Howell, et al. (1997). The Jeans Condition: A New Constraint on Spatial Resolution in Simulations of Isothermal Self-gravitational Hydrodynamics. Astrophysical Journal Letters 489, L179–L183.

Tytler, D., P. Paschos, D. Kirkman, M. Norman, and T. Jena (2009). The effect of large-scale power on simulated spectra of the lya forest. Monthly Notices of the Royal Astronomical Society 393, 723 – 758.

Upadhye, A., J. Kwan, A. Pope, K. Heitmann, S. Habib, et al. (2016). Redshift-space distortions in massive neutrino and evolving dark energy cosmologies. Phys. Rev. D 93(6), 063515. `arXiv:astro-ph.CO/1506.07526`.

van der Wel, A., M. Franx, P. G. van Dokkum, R. E. Skelton, I. G. Momcheva, et al. (2014). 3D-HST+CANDELS: The Evolution of the Galaxy Size-Mass Distribution since z = 3. Astrophysical Journal 788(1), 28. `arXiv:astro-ph.GA/1404.2844`.

van Dokkum, P. G. and G. Brammer (2010). Hubble Space Telescope WFC3 Grism Spectroscopy and Imaging of a Growing Compact Galaxy at z = 1.9. Astrophysical Journal Letters 718(2), L73–L77. `arXiv:astro-ph.CO/1003.3446`.

van Leer, B. (1979). Towards the Ultimate Conservative Difference Scheme. V. A Second-Order Sequel to Godunov's Method. Journal of Computational Physics 32(1), 101–136.

Viel, M., G. D. Becker, J. S. Bolton, and M. G. Haehnelt (2013). Warm dark matter as a solution to the small scale crisis: New constraints from high redshift lyman-αforest data. Physical Review D 88(4).

Viel, M., G. D. Becker, J. S. Bolton, M. G. Haehnelt, M. Rauch, et al. (2008). How cold is cold dark matter? small-scales constraints from the flux power spectrum of the high-redshift lyman-αforest. Physical Review Letters 100(4).

Viel, M., M. G. Haehnelt, and V. Springel (2004). Inferring the dark matter power spectrum from the lymanαforest in high-resolution qso absorption spectra. Monthly Notices of the Royal Astronomical Society 354(3), 684–694.

Viel, M., M. G. Haehnelt, and V. Springel (2006). Testing the accuracy of the hydrodynamic particle-mesh approximation in numerical simulations of the Lyman α forest. MNRAS 367(4), 1655–1665. `arXiv:astro-ph/astro-ph/0504641`.

Viel, M., M. G. Haehnelt, and V. Springel (2010). The effect of neutrinos on the matter distribution as probed by the intergalactic medium. J. Cosmology Astropart. Phys. 2010(6), 015. `arXiv:astro-ph.CO/1003.2422`.

Viel, M., J. Lesgourgues, M. G. Haehnelt, S. Matarrese, and A. Riotto (2005). Constraining warm dark matter candidates including sterile neutrinos and light gravitinos with WMAP and the Lyman-α forest. Phys. Rev. D 71(6), 063534. `arXiv:astro-ph/astro-ph/0501562`.

Villaescusa-Navarro, F., F. Marulli, M. Viel, E. Branchini, E. Castorina, et al. (2014). Cosmology with massive neutrinos I: towards a realistic modeling of the relation between matter, haloes and galaxies. J. Cosmology Astropart. Phys. 2014(3), 011. `arXiv:astro-ph.CO/1311.0866`.

Vogelsberger, M., J. Zavala, and A. Loeb (2012). Subhaloes in self-interacting galactic dark matter haloes. MNRAS 423(4), 3740–3752. arXiv:astro-ph.CO/1201.5892.

Volonteri, M., Y. Dubois, C. Pichon, and J. Devriendt (2016). The cosmic evolution of massive black holes in the Horizon-AGN simulation. MNRAS 460(3), 2979–2996. arXiv:astro-ph.GA/1602.01941.

Walther, M., J. F. Hennawi, H. Hiss, J. Oñorbe, K.-G. Lee, et al. (2018). A New Precision Measurement of the Small-scale Line-of-sight Power Spectrum of the Lyα Forest. Astrophysical Journal 852, 22. arXiv:1709.07354.

Wang, J. and S. D. M. White (2007). Discreteness effects in simulations of hot/warm dark matter. MNRAS 380(1), 93–103. arXiv:astro-ph/astro-ph/0702575.

Wang, M.-Y., R. A. C. Croft, A. H. G. Peter, A. R. Zentner, and C. W. Purcell (2013). Lyman-α forest constraints on decaying dark matter. Phys. Rev. D 88(12), 123515. arXiv:astro-ph.CO/1309.7354.

Wang, T., C. Schreiber, D. Elbaz, Y. Yoshimura, K. Kohno, et al. (2019). A dominant population of optically invisible massive galaxies in the early Universe. Nature 572(7768), 211–214. arXiv:astro-ph.GA/1908.02372.

Weinberg, D. H., J. Miralda-Escudé, L. Hernquist, and N. Katz (1997). A Lower Bound on the Cosmic Baryon Density. Astrophysical Journal 490(2), 564–570. arXiv:astro-ph/astro-ph/9701012.

Weinberg, S. (1989). The cosmological constant problem. Reviews of Modern Physics 61(1), 1–23.

Weinberger, R., V. Springel, R. Pakmor, D. Nelson, S. Genel, et al. (2018). Supermassive black holes and their feedback effects in the illustristng simulation. Monthly Notices of the Royal Astronomical Society 479(3), 4056–4072.

Weinheimer, C. and KATR. I. N. Collaboration (2002). KATRIN, a next generation tritium β decay experiment in search for the absolute neutrino mass scale. Progress in Particle and Nuclear Physics 48(1), 141–150.

Weinmann, S. M., A. Pasquali, B. D. Oppenheimer, K. Finlator, J. T. Mendel, et al. (2012). A fundamental problem in our understanding of low-mass galaxy evolution. MNRAS 426(4), 2797–2812. arXiv:astro-ph.CO/1204.4184.

Welker, C., Y. Dubois, J. Devriendt, C. Pichon, S. Kaviraj, et al. (2017). The rise and fall of stellar discs across the peak of cosmic star formation history: effects of mergers versus diffuse stellar mass acquisition. MNRAS 465(1), 1241–1258. arXiv:astro-ph.GA/1502.05053.

Weymann, R. J., S. L. Morris, C. B. Foltz, and P. C. Hewett (1991). Comparisons of the Emission-Line and Continuum Properties of Broad Absorption Line and Normal Quasi-stellar Objects. Astrophysical Journal 373, 23.

White, M. (1999). Anisotropy in the microwave background. In D. O. Caldwell (Ed.), COSMO-98, Volume 478 of American Institute of Physics Conference Series, pp. 157–163.

White, S. D. M. and M. J. Rees (1978). Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering. MNRAS 183, 341–358.

Williams, C. C., M. Giavalisco, P. Cassata, E. Tundo, T. Wiklind, et al. (2014). The Progenitors of the Compact Early-type Galaxies at High Redshift. Astrophysical Journal 780(1), 1. arXiv:astro-ph.CO/1310.3819.

Wright, A. H. et al. (2018). KiDS+VIKING-450: A new combined optical & near-IR dataset for cosmology and astrophysics. arXiv:astro-ph.CO/1812.06077.

Xu, W. L., C. Dvorkin, and A. Chael (2018). Probing sub-GeV dark matter-baryon scattering with cosmological observables. Phys. Rev. D 97(10), 103530. arXiv:astro-ph.CO/1802.06788.

Yèche, C., N. Palanque-Delabrouille, J. Baur, and H. d. M. d. Bourboux (2017). Constraints on neutrino masses from lyman-alpha forest power spectrum with boss and xq-100. Journal of Cosmology and Astroparticle Physics 2017(06), 047–047.

Yepes, G., R. Kates, A. Khokhlov, and A. Klypin (1997). Hydrodynamical simulations of galaxy formation: effects of supernova feedback. MNRAS 284(1), 235–256. arXiv:astro-ph/astro-ph/9605182.

York, D. G., J. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, et al. (2000). The sloan digital sky survey: Technical summary. The Astronomical Journal 120(3), 1579–1587.

Yoshikawa, K., N. Yoshida, and M. Umemura (2013). Direct Integration of the Collisionless Boltzmann Equation in Six-dimensional Phase Space: Self-gravitating Systems. Astrophysical Journal 762(2), 116. arXiv:astro-ph.IM/1206.6152.

Zaldarriaga, M. (1997). Polarization of the microwave background in reionized models. Physical Review D 55, 1822–1829.

Zel'dovich, Y. B. (1964). Special Issue: the Theory of the Expanding Universe as Originated by a. a. Fridman. Soviet Physics Uspekhi 6(4), 475–494.

Zel'Dovich, Y. B. (1970). Reprint of 1970A&A.....5...84Z. Gravitational instability: an approximate theory for large density perturbations. Astronomy & Astrophysics 500, 13–18.

Zhang, J., H. Liu, and M.-C. Chu (2018). Cosmological Simulation for Fuzzy Dark Matter Model. Frontiers in Astronomy and Space Sciences 5, 48. arXiv:astro-ph.CO/1809.09848.

Zhang, Y., P. Anninos, and M. L. Norman (1995). A multispecies model for hydrogen and helium absorbers in lyman-alpha forest clouds. The Astrophysical Journal 453(2).

Zhao, G.-B., Y. Wang, A. J. Ross, S. Shand era, W. J. Percival, et al. (2016). The extended Baryon Oscillation Spectroscopic Survey: a cosmological forecast. MNRAS 457(3), 2377–2390. arXiv:astro-ph.CO/1510.08216.

Zolotov, A., A. Dekel, N. Mandelker, D. Tweed, S. Inoue, et al. (2015). Compaction and quenching of high-z galaxies in cosmological simulations: blue and red nuggets. MNRAS 450(3), 2327–2353. arXiv:astro-ph.GA/1412.4783.

Zuntz, J. et al. (2018). Dark Energy Survey Year 1 Results: Weak Lensing Shape Catalogues. Mon. Not. Roy. Astron. Soc. 481(1), 1149–1182. `arXiv:astro-ph.CO/1708.01533`.

Zuo, L. and J. R. Bond (1994). The Transmission Correlation in the QSO LY alpha Forest Produced by Finite Width Lines. Astrophysical Journal 423, 73.

Zwicky, F. (1933). Die Rotverschiebung von extragalaktischen Nebeln. Helvetica Physica Acta 6, 110–127.

université
PARIS-SACLAY

ÉCOLE DOCTORALE
Particules, hadrons, énergie et noyau :
instrumentation, imagerie, cosmos
et simulation (PHENIICS)

Titre: Neutrinos et matière noire à la lumière des quasars

Mots clés: Neutrinos, Matière noire, Lyman-alpha, Physique des baryons, Simulations numériques, Relevé spectroscopique

Résumé: Bien que le modèle cosmologique standard ΛCDM offre un cadre remarquablement en accord avec de nombreuses observations indépendantes, beaucoup de mystères persistent. En particulier, la masse des neutrinos est toujours inconnue. De par leur masse non nulle, ils laissent une empreinte sur les grandes structures de l'univers à travers l'échelle à laquelle ils diffusent, qui se manifeste comme un déficit de fluctuations de densité de matière aux petites échelles. Par ailleurs, le modèle de matière noire froide montre des tensions persistantes avec les observations sur des échelles inférieures au Mpc. Dans ce travail de thèse, j'utilise le spectre de puissance de la transmission de flux dans la forêt Lyman-α dans des spectres de quasars lointains afin de contraindre la somme des masses des neutrinos et étudier la plausibilité d'un modèle de matière noire tiède. Dans un premier temps, j'ai mesuré le spectre de puissance 1D de la forêt Lyα à 13 redshifts tel que $2 \leq z \leq 4.25$ en utilisant 43,751 spectres de quasars de haute qualité des programmes BOSS et eBOSS du relevé spectroscopique SDSS. Afin d'obtenir des résultats robustes sachant que les incertitudes statistiques ont été réduites d'un facteur 3 par rapport à la mesure précédente, j'ai identifié et controlé plusieurs effets systématiques dans l'analyse de données. Modéliser le flux Lyα implique de recourir à des simulations hydrodynamiques cosmologiques sachant qu'il émerge de la combinaison complexe entre l'évolution des grandes structures et de la physique baryonique des petites échelles. En effet, les processus astrophysiques tels que la formation stellaire et les phénomènes de retro-action de supernovae et des trous noirs rejettent une quantité considérable d'énergie dans le milieu environnant et modifient l'état thermique et la distribution du gaz dans le milieu intergalactique. Afin d'améliorer les prédictions théoriques de la forêt Lyα à un niveau comparable des données, je contrains l'impact des mécanismes de retro-action des trous noirs sur le spectre de puissance Lyα grâce à un jeu de 8 simulations que j'ai produit en me basant sur des observations astrophysiques et qui couvre l'ensemble des modèles de feedback plausibles. Je fournis une correction analytique de cet effet ainsi qu'une limite supérieure et inférieure tel que $2 \leq z \leq 4.25$ et je montre qu'ignorer un tel mécanisme induit un biais de 2σ sur n_s et 1σ sur σ_8. Finalement, je combine les mesure des spectres de puissance Lyα avec des données CMB et BAO afin de les comparer statistiquement aux prédictions théoriques des simulations hydrodynamiques pour améliorer la contrainte sur la somme des masses des neutrinos de $\sum m_\nu < 0.12$ eV établie précédemment à $\sum m_\nu < 0.09$ eV dans le cas le plus extrême à 95% de vraisemblance. Ce résultat indique une préférence des données pour un modèle normal de hiérarchie de masse. La combinaison des données Lyα eBOSS avec les données LyαXQ-100, contraint la masse des reliques thermiques à $m_X > 5.3$ keV à 95% de vraisemblance dans un cas de modèle de matière noire entièrement constitué de matière noire tiède, ce qui se traduit par une limite supérieure sur les neutrinos stériles produits de manière non-résonante à $m_{\nu_s} > 34$ kev. Aussi, les données Lyα-eBOSS confirme une tension existante avec les données CMB sur l'indice spectral n_s et indique une préférence pour une dépendance d'échelle non nulle de n_s à 3σ.

Title: Neutrinos and dark matter cosmology with the Lyman-α forest: the interplay between large-scale evolution and small-scale baryonic physics

Keywords: Neutrinos, Dark Matter, Lyman-alpha, Baryonic physics, Numerical simulations, Spectroscopic survey

Abstract: Even if the standard cosmological ΛCDM model provides a remarkably successful framework to explain many independent observations, it sill faces many challenges. In particular, the masses of neutrinos are still unknown and significantly alter structure formation because of their free-streaming that suppresses density fluctuations below a typical length scale inversely proportional to their rest mass. In addition, the cold dark matter (CDM) scenario is in tension with observations on scales smaller than the Mpc. In this thesis work, I use the power spectrum of the transmitted flux in the Lyman-α (Lyα) forest of distant quasar spectra to constrain the sum of neutrino masses, $\sum m_\nu$, and determine the plausibility of a warm dark matter model, which is conveniently consistent with cold dark matter predictions on large scales while circumventing its issues at small scales because of its non-negligible velocity dispersion. First I measure the 1D power spectrum of the Lyα forest of 43,751 high quality quasar spectra between $2 \leq z \leq 4.6$ from the BOSS and eBOSS programs of the SDSS spectroscopic survey. To obtain robust results given the unprecedented statistical power of the data I perform a careful investigation of observational systematic sources and their sources. Modeling the Lyα flux power spectrum requires to run hydrodynamical cosmological simulations because it arises from the complex interplay between large-scale structure evolution and small-scale baryonic physics. Indeed, astrophysical processes such as star formation or AGN feedback inject energy in the ambient medium and strongly impact the thermal state and gas distribution in the intergalactic medium. Including such processes in hydrodynamical simulations requires to rely on arbitrary parameters calibrated on astrophysical observations leading to discrepancies between different state-of-the-art simulations. In order to improve theoretical predictions of the Lyα forest, I constrain the impact of AGN feedback using a series of 8 hydro-cosmological simulations covering the whole plausible range of feedback models. I provide upper and lower limit for this signature for $2 \leq z \leq 4.25$ and also show that ignoring this effect leads to 2σ shift on n_s and 1σ shift on σ_8. Finally, I combine the Lyα flux power spectrum measurements with CMB data, BAO data and theoretical predictions from hydrodynamical simulations to enhance the previously established constraints on the sum of neutrino masses from $\sum m_\nu < 0.12$ eV to the most stringent constraints to date $\sum m_\nu < 0.09$ eV in the most extreme case with 95% confidence, which tends to favor the normal hierarchy neutrino mass scenario. Combining eBOSS with XQ-100 Lyα data the mass m_X of hypothetical thermal relics is constrained to $m_X > 5.3$ keV at the 95% confidence level in the case of a pure warm dark matter scenario, which translates into $m_{\nu_s} > 34$ kev for non-resonantly produced sterile neutrinos. Also, a mild-tension is found on n_s between eBOSS Lyα and CMB data, which translates into a preference for a non-zero running of n_s at the level of about 3σ.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France