



**HAL**  
open science

# Large surveys: from galaxy evolution to cosmological probes

Louis Legrand

► **To cite this version:**

Louis Legrand. Large surveys: from galaxy evolution to cosmological probes. Cosmology and Extra-Galactic Astrophysics [astro-ph.CO]. Université Paris-Saclay, 2020. English. NNT : 2020UPASP023 . tel-03164609

**HAL Id: tel-03164609**

**<https://theses.hal.science/tel-03164609v1>**

Submitted on 10 Mar 2021

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

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

# Large surveys: from galaxy evolution to cosmological probes

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

École doctorale n°127, Astronomie et Astrophysique  
d'Île-de-France (AAIF)  
Spécialité de doctorat : Astronomie et Astrophysique  
Unité de recherche : Université Paris-Saclay, CNRS, Institut  
d'astrophysique spatiale, 91405, Orsay, France.  
Réfèrent : Faculté des sciences d'Orsay

**Thèse présentée et soutenue à Orsay, le 29/09/2020, par**

**Louis Legrand**

## Composition du jury:

<b>Martin Kilbinger</b> Directeur de recherche, CEA Paris-Saclay	Président
<b>Carlo Baccigalupi</b> Professeur, SISSA, Trieste, Italie	Rapporteur
<b>Stéphanie Escoffier</b> Directrice de recherche, CPPM, Aix-Marseille Université	Rapporteur
<b>Jean-Christophe Hamilton</b> Directeur de recherche, APC, Université de Paris	Examineur
<b>Sylvain de la Torre</b> Astronome, LAM, Aix-Marseille Université	Examineur
<b>Marian Douspis</b> Astronome, IAS, Orsay	Directeur de thèse
<b>Nabila Aghanim</b> Directrice de recherche, IAS, Orsay	Co-directrice de thèse



*À mes amis, ma famille, et celle qui se reconnaîtra.*



*Le silence éternel de ces espaces infinis m'effraie.*  
Pensées – Pascal

*Adso, dit Guillaume, résoudre un mystère n'est pas la même chose qu'une déduction à partir de principes premiers. Et ça n'équivaut pas non plus à recueillir une bonne quantité de données particulières pour inférer ensuite une loi générale. Cela signifie plutôt se trouver en face d'une, ou deux, ou trois données particulières qui apparemment n'ont rien en commun, et chercher à imaginer si elles peuvent être autant de cas d'une loi générale que tu ne connais pas encore, et qui n'a peut-être jamais été énoncée.*  
Le nom de la Rose – Umberto Ecco



# Remerciements

Quand je repense à ces trois années passées à l'IAS, il m'est difficile d'isoler chaque événement, ainsi que de définir précisément l'évolution que j'ai connue. Il s'est passé tellement de choses dans un laps de temps si compact que tout semble uniformisé. Comme il m'est impossible de résumer tous mes souvenirs ici, je vais me contenter de remercier les personnes que j'ai pu rencontrer et qui m'ont soutenues, en espérant n'oublier personne.

Avant de rentrer dans des détails plus personnels, je souhaite d'abord remercier chaleureusement les membres de mon jury pour s'être intéressés à mon travail, pour avoir lu avec attention mon manuscrit et pour leurs questions et commentaires pertinents lors de ma soutenance.

Marian, merci à toi pour m'avoir suivi et encouragé tout au long de ma thèse, pour ta patience et pour tes précieux conseils. Merci à toi aussi Nabila pour avoir cru en moi et m'avoir poussé à toujours me dépasser sans me reposer sur mes acquis. Marian tu as été mon guide qui m'a permis de garder l'équilibre, tandis que Nabila me poussait sans cesse en dehors de ma zone de confort pour me faire avancer.

Un énorme merci à Mathieu, Hervé et Julien et à l'équipe de cosmo pour m'avoir accueilli et pour m'avoir donné envie de faire de la recherche lorsque que j'étais un jeune stagiaire. Merci pour votre patience lorsque vous répondiez à mes questions sur beaucoup de sujets divers et variés.

Merci mille fois aux thésards et post-docs avec qui j'ai passé des moments mémorables et de partagés de bonnes tranches de rigolades, merci Nadège, Laura, Giulio, Édouard, Daniela, Thibaut, Thomas, Antoine, Thiébaud, Natalia, Joseph, Nicola, Stéphane, Fabien, Tony, Marion, Hideki, Céline.

Un coucou tout particulier à Victor et Adélie, avec qui nous formions les stagiaires du dimanche, puis les thésards du dimanche et maintenant les post-docs du dimanche. La suite au prochain épisode . . .

Merci à toute l'équipe de direction de l'IAS pour m'avoir accueilli dans le laboratoire et pour m'avoir donné les moyens de travailler sereinement.

Merci Henry pour m'avoir offert un super stage et pour les cafés dans ton bureau.

Merci à Carlos et Guillaume pour m'avoir accueillis à Teruel, ce qui fut un véritable tournant dans ma thèse.

Merci aux camarades rencontrés lors des écoles d'été Euclid, la liste est trop longue pour tous vous ajouter ici. Merci pour ces très bons moments passés avec vous dans les bars de Roscoff et sur les plages de Banyuls.

Merci à mes amis de longue date, Romain, Seb, Gus qui sont avec moi depuis longtemps (15 ans !) et qui m'ont toujours suivis et soutenus à distance.

Bien sûr merci à Camille, Basile, Grégoire, Bénédicte, Valentin, Augustin, vous

## REMERCIEMENTS

---

êtes toujours là pour moi et franchement c'était super de vous avoir tout au long de ma thèse.

Et enfin, un énorme merci à mes parents, mes grands parents, mes oncles et tantes, mes cousins, pour m'avoir toujours encouragés à faire ce que je souhaitais, pour m'avoir laissé une grande liberté et pour m'avoir toujours soutenu en toutes circonstances.

# Abstract

Large galaxy surveys are like open windows on our Universe: they provide precious insights on its components and on its evolution. On the one hand, pencil surveys go deep into the cosmos to explore the formation and evolution of galaxies. On the other hand, wide surveys are mapping the distribution of matter on large scales to understand the nature of dark energy and dark matter.

During my thesis, I explore the capabilities of these large surveys to address the following questions:

1. What are the main drivers of galaxy evolution? More specifically, I investigate the relation between the dark matter halo and the host galaxies and I explore how galaxies build their stellar content.
2. What improvements on our knowledge of the Universe will be brought by upcoming wide galaxy surveys? To maximize this improvements, I use an alternative galaxy probe which helps in breaking degeneracies with standard probes. I then investigate the gain of combining galaxy surveys with next generation cosmic microwave background (CMB) surveys.

Using precise galaxy stellar-mass function measurements in the COSMOS field, I first determine the stellar-to-halo mass relation through a parametric abundance matching technique (Legrand et al. 2019). Thanks to the completeness of the COSMOS survey from  $z \sim 0.2$  to  $z \sim 5$ , I obtain for the first time this relation over such a large redshift range from a single coherent sample. I find that the ratio of stellar-to-halo mass content peaks at a characteristic halo mass which increases up to  $z = 2.3$  and remains flat up to  $z = 4$ . This steady increase of the characteristic halo mass questions the role of cold gas inflows as drivers of galaxy formation at high redshift. To address this topic, I link observations of the cold molecular gas content in galaxies up to  $z = 4$  to the evolution of the dark matter halo mass. I find that the joint evolution of cold gas mass fraction and halo mass is in agreement with the hypothesis of cold gas inflows being responsible of efficient galaxy formation at high redshift.

With the scope of maximising the cosmological power of next generation spectroscopic surveys, I show that a novel cosmological observable, the angular redshift fluctuations (ARF) will provide complementary cosmological information in addition to the standard angular galaxy clustering (Legrand et al. 2020). Due to its distinct sensitivity to the peculiar velocity field, I find that the cosmological and galaxy bias parameters express different degeneracies when inferred from ARF or from angular galaxy clustering. As such, combining both observables breaks these

degeneracies and greatly decreases the marginalised uncertainties, by a factor of at least two on most parameters in the  $\Lambda$ CDM and wCDM models.

As part of the *Euclid* collaboration, I then investigate the cosmological power of the upcoming *Euclid* survey, which will offer us an exquisite measurement of the matter distributions on the full extra-galactic sky. In order to fully exploit all the potential of the *Euclid* survey it is crucial to combine it with upcoming CMB surveys. I use the Fisher formalism to forecast the benefits of performing a joint analysis of CMB probes with *Euclid* main probes (weak lensing and galaxy clustering) ([Euclid Collaboration in prep.](#)). I test both the standard cosmological model,  $\Lambda$ CDM, and its extensions, and show that CMB will improve the constraints by a factor two on most cosmological parameters, and most notably on dark energy modified models which are of key interest for *Euclid*.

# Résumé

Les grands relevés de galaxies sont des fenêtres ouvertes sur notre Univers: ils nous offrent de précieuses informations sur son contenu et sur son évolution. D'une part les relevés profonds explorent la formation et l'évolution des galaxies. D'autre part, les relevés à grand champ cartographient la distribution de la matière dans le but de comprendre la nature de l'énergie noire et de la matière noire.

Au cours de cette thèse, j'explore les capacités offertes par ces relevés afin de répondre aux questions suivantes:

1. Quels sont les principaux moteurs de l'évolution des galaxies ? Plus précisément, j'étudie la relation entre les galaxies et leurs halos de matière noire, et j'explore leur efficacité à produire des étoiles.
2. Quels progrès dans notre connaissance de l'Univers seront apportés par les futurs relevés de galaxies ? Dans le but de maximiser les résultats de ces relevés, j'utilise une nouvelle sonde cosmologique qui permet, en combinaison avec les sondes standards, de briser les dégénérescences entre les paramètres du modèle cosmologique. J'analyse ensuite le gain apporté par la combinaison des relevés de galaxies avec des relevés du fond diffus cosmologique (CMB) de prochaine génération.

Je commence par déterminer la relation entre la masse stellaire et la masse des halos de matière noire des galaxies en utilisant des mesures précises de la fonction de masse stellaire dans le champ COSMOS (Legrand et al. 2019). Grâce à l'exhaustivité du relevé COSMOS entre  $z \sim 0.2$  et  $z \sim 5$ , j'obtiens pour la première fois cette relation sur une aussi grande gamme de redshifts à partir d'un seul relevé. Je constate que la masse de halo caractéristique, définie comme maximisant le rapport entre la masse stellaire et la masse du halo, augmente entre  $z = 0$  et  $z = 2.3$  et reste stable jusqu'à  $z = 4$ . Cette augmentation de la masse de halo caractéristique met en lumière le rôle des flux de gaz froid comme moteurs de la formation des galaxies à grand redshift. Afin d'approfondir ce sujet, je combine des observations de la teneur en gaz moléculaire froid des galaxies jusqu'à  $z = 4$ , avec la relation entre masse stellaire et masse du halo de matière noire. Je constate que l'évolution de la fraction de masse du gaz froid est en accord avec l'hypothèse selon laquelle les apports de gaz froid sont responsables de la plus grande efficacité de formation des galaxies à grand redshift dans les halos massifs.

Ensuite, dans le but de maximiser les contraintes cosmologiques qui seront apportées par les prochains grands relevés spectroscopiques, je montre qu'une nouvelle observable, les fluctuations angulaires de redshift (ARF), apportent des informations

complémentaires par rapport au traditionnel *angular galaxy clustering* (Legrand et al. 2020). Grâce à la sensibilité particulière des ARF au champ de vitesse de la matière, je montre que les dégénérescences entre les paramètres cosmologiques et de biais des galaxies sont différentes lorsqu'elles se basent sur les ARF ou sur le *angular galaxy clustering*. Dès lors, la combinaison des deux observables permet de lever des dégénérescences et d'améliorer les contraintes, d'un facteur au moins deux, sur la plupart des paramètres des modèles  $\Lambda$ CDM et wCDM.

Finalement, en tant que membre de la collaboration *Euclid*, j'ai exploré le potentiel cosmologique de ce futur relevé de galaxies. Ce relevé nous permettra de mesurer très précisément la distribution de la matière sur tout le ciel extra-galactique. Dans le but d'exploiter entièrement tout son potentiel, il est crucial de le combiner avec les futurs relevés du CMB. J'utilise le formalisme de Fisher afin de prédire l'intérêt d'une analyse combinée des sondes CMB avec les sondes *Euclid* (Euclid Collaboration in prep.). Je teste à la fois le modèle  $\Lambda$ CDM standard et ses extensions, et montre que le CMB améliorera les contraintes d'un facteur au moins deux sur la plupart des paramètres cosmologiques, et notamment sur les modèles d'énergie noire alternatifs, qui font partie des intérêts majeurs pour *Euclid*.

# Synthèse en Français

Dans le but de découvrir la nature de l'énergie sombre et de la matière sombre, de nouveaux relevés à grand champ, tels que Euclid, DESI et LSST, vont cartographier la distribution des galaxies dans l'Univers, au cours de la prochaine décennie. Ces relevés utiliseront les galaxies comme traceurs de la distribution de la matière. Celles-ci se forment et évoluent au sein de halos de matière noire. Cependant nous savons que les galaxies sont un traceur biaisé de la matière: la densité de galaxies est plus grande que la densité de matière dans les zones denses, et est plus faible dans les zones sous-denses. De plus, aux petites échelles les effets complexes de la physique des baryons tels que les noyaux actifs de galaxies vont empêcher le gaz de refroidir et de former des étoiles, et vont avoir un impact sur la distribution de la matière. Ainsi, si nous souhaitons maximiser les informations apportées par les futurs grands relevés de galaxies, il est important de comprendre précisément les processus de la formation des galaxies et la relation avec leurs halos de matière noire.

La première partie de ma thèse se concentre sur cette problématique. Plus précisément, je commence par déterminer la relation entre la masse stellaire et la masse des halos de matière noire des galaxies en utilisant des mesures précises de la fonction de masse stellaire dans le champ COSMOS (Legrand et al. 2019). Grâce à l'exhaustivité du relevé COSMOS entre  $z \sim 0.2$  et  $z \sim 5$ , j'obtiens pour la première fois cette relation sur une aussi grande gamme de redshifts à partir d'un seul relevé. Comme le montre la Figure 1, je constate que la masse de halo caractéristique, définie comme maximisant le rapport entre la masse stellaire et la masse du halo, augmente entre  $z = 0$  et  $z = 2.3$  et reste stable jusqu'à  $z = 4$ . Cette augmentation de la masse de halo caractéristique met en lumière le rôle des flux de gaz froid comme moteurs de la formation des galaxies à grand redshift. Afin d'approfondir ce sujet, je combine des observations de la teneur en gaz moléculaire froid des galaxies jusqu'à  $z = 4$ , avec la relation entre masse stellaire et masse du halo de matière noire. Je constate que l'évolution de la fraction de masse du gaz froid est en accord avec l'hypothèse selon laquelle les apports de gaz froid sont responsables de la plus grande efficacité de formation des galaxies à grand redshift dans les halos massifs. Pour aller plus loin, il sera nécessaire d'obtenir des observations directes de filament de gaz froid nourrissant le coeur des halos massifs à grand redshift.

Une autre source majeure d'informations sur la distribution de la matière est le fond diffus cosmologique (CMB). Comme les photons du CMB ont traversé tout l'Univers pour nous parvenir, leur chemin a été dévié par les effets de lentillage gravitationnel. Reconstruire cette déviation permet d'obtenir un traceur non biaisé de la distribution de la matière. Dans la prochaine décennie, de nouvelles observations du CMB telles que Advanced ACT, SPT-3G et le Simons Observatory permettront

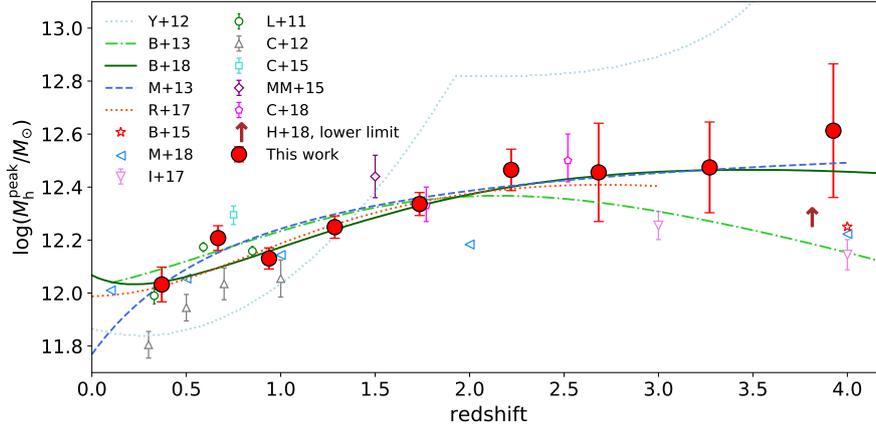


Figure 1: Masse des halos maximisant le ratio masse stellaire sur masse de matière noire, en fonction du redshift (points rouges). Nous montrons ici des exemples de la littérature pour comparaison. Grâce au relevé COSMIS, nos résultats sont les premiers à être issus d’un échantillon cohérent sur un aussi grand interval de redshift. Ceci montre que les halos massifs étaient plus efficaces pour créer des étoiles à grand redshift qu’à bas redshift.

d’améliorer la résolution aux petites échelles. Combiner ce traceur avec la distribution des galaxies permet de lever les dégénérescences dû au biais des galaxies.

En tant que membre de la collaboration *Euclid*, j’ai exploré le potentiel cosmologique de l’analyse combinée de ce futur relevé de galaxies avec les futures observations du CMB. J’utilise le formalisme de Fisher afin de prédire l’intérêt d’une analyse combinée des sondes CMB avec les sondes *Euclid* ([Euclid Collaboration in prep.](#)). Je teste à la fois le modèle  $\Lambda$ CDM standard et ses extensions, et montre que le CMB améliorera les contraintes d’un facteur au moins deux sur la plupart des paramètres cosmologiques, et notamment sur les modèles d’énergie noire alternatifs, qui font partie des intérêts majeurs pour *Euclid*.

Ensuite, dans le but de maximiser les contraintes cosmologiques qui seront apportées par les prochains grands relevés spectroscopiques, je montre qu’une nouvelle observable, les fluctuations angulaires de redshift (ARF), apportent des informations complémentaires par rapport au traditionnel *angular galaxy clustering* ([Legrand et al. 2020](#)). Grâce à la sensibilité particulière des ARF au champ de vitesse de la matière, je montre que les dégénérescences entre les paramètres cosmologiques et de biais des galaxies sont différentes lorsqu’elles se basent sur les ARF ou sur le *angular galaxy clustering*. Dès lors, la combinaison des deux observables permet de lever des dégénérescences et d’améliorer les contraintes, d’un facteur au moins deux, sur la plupart des paramètres des modèles  $\Lambda$ CDM et  $w$ CDM, comme le montre la Figure 2. Ces résultats sont très prometteurs, et je souhaite poursuivre le développement des ARF en incluant la masse des neutrinos et les modèles de gravité modifiée dans cette analyse.

En conclusion, afin de maximiser le retour d’information des futures observations de l’Univers, il sera nécessaire de combiner les relevés de galaxies et les relevés du CMB. De plus, il sera nécessaire de développer le formalisme permettant d’extraire

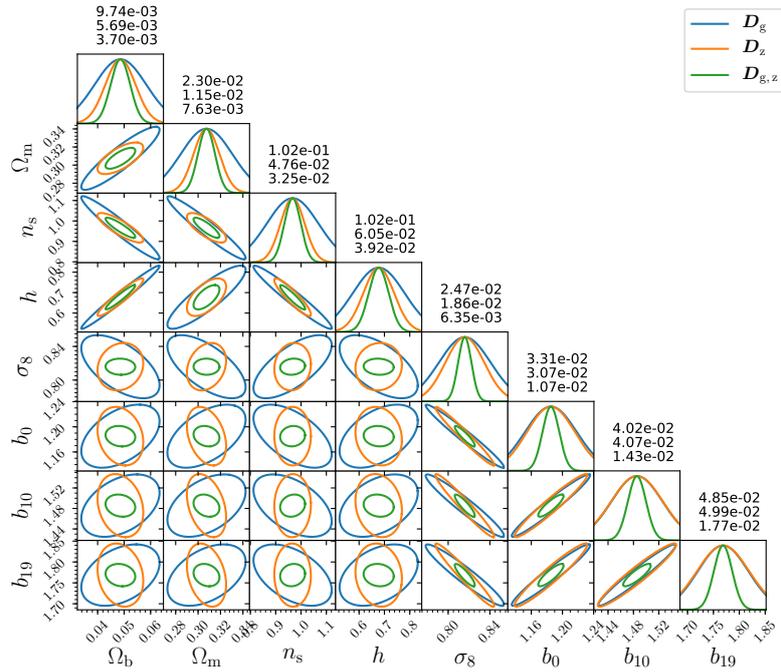


Figure 2: Contraintes à  $1\sigma$  pour 5 paramètres cosmologiques et trois paramètres de biais des galaxies, pour un relevé de type DESI. Les lignes bleues donnent les contraintes pour les *angular galaxy clustering*, les lignes oranges pour les ARF et les lignes vertes sont la combinaison des deux. Nous voyons que les ARF et les *angular galaxy clustering* ont des dégénérescences différentes qui sont brisées en combinant les deux.

toute l'information contenue aux petites échelles, que ce soit dans la distribution des galaxies, en tenant compte des effets de la physique des baryons, ou dans les estimateurs du lentillage du CMB, qui ne sont pour l'instant pas optimaux pour les petites échelles. Ce sont ces deux axes de recherche, physique des baryons et estimateurs du lentillage du CMB, que je souhaite développer afin de poursuivre les travaux présentés dans cette thèse.

# Introduction

Even if the standard model of cosmology has seen great successes in the last two decades, its two main components, dark matter and dark energy, remain a mystery. Indeed, this model does not provide any clue about their inner nature. Understanding this dark sector is the main goal for the next decade in cosmology.

Within a few years, wide galaxy surveys such as DESI, *Euclid*, LSST and WFIRST will measure the distribution of matter on the full extra-galactic sky with exquisite precision. Using both galaxy distribution and the weak gravitational lensing as cosmological probes, these wide surveys will test gravity at cosmological scales and investigate the nature of the accelerated expansion of the Universe.

These surveys will use galaxies as a tracer of the (mostly invisible) underlying distribution of matter. Galaxies are indeed tightly interwoven with dark matter: galaxies form and evolve inside virialised collapsed structures of dark matter known as haloes. However galaxies are a biased tracer of matter, they do not sample the density field uniformly: the density of galaxies is higher than the matter density in dense areas, and lower in underdense areas. Moreover at the smallest scales, baryonic effects, such as active galactic nuclei feedback, will prevent star formation and redistribute baryonic material on larger scales. Even if the broad picture of the galaxy formation and evolution is known, the exact physical mechanisms driving these processes remain to be elucidated. As such, if we want to maximize the cosmological power of future galaxy surveys, and understand the relation between galaxies and dark matter, we first have to understand precisely the processes of galaxy formation.

The first part of my PhD thesis focus on this problematic. More specifically, I measure the evolution of the stellar mass to halo mass relation of galaxies. In order to test a scenario explaining the evolution I observed, where cold gas inflows plays an important role at high redshift, I then measure the joint evolution of the cold molecular gas with the halo mass of galaxies.

Another major source of information on our Universe is the cosmic microwave background (CMB). In the last two decades, precise observations of its tiny temperature fluctuations allowed us to get a picture of the primordial Universe. Moreover, because the photons of the CMB interacted with all the structures along their path, they carry precious information about the composition and the evolution of the Universe. Ongoing or upcoming experiments such as Advanced ACT, SPT-3G, the Simons Observatory and later on the CMB-Stage 4 will improve the resolution at small scales compared to *Planck*, which will allow to accurately map the distribution of matter in the Universe. This coincidence in the acquisition of datasets of tremendous quality and huge size from both the large scale structures and CMB fronts

implies that we must develop the tools to fully exploit all their potential. Indeed, combining these galaxy and CMB surveys will allow us to identify and mitigate all the possible systematics, such as the galaxy bias discussed above, and validate the results we obtain. But we first need to develop the tools to predict the cosmological power for a combined analysis of these next generation surveys.

The second part of my PhD thesis presents the research performed inside the *Euclid* XCMB science working group, which is dedicated to combine *Euclid* and CMB observables. I show how combining the *Euclid* survey with next generation CMB surveys will be of major interest to reach the high accuracy required to test the nature of the dark sector.

Moreover, to go further in the exploitation of next generation surveys, alternative cosmological probes are being developed. The interest of these new probes must be tested in order to validate their use in cosmological analyses. The last part of my manuscript, is dedicated to this work. I introduce an alternative probe which uses the fluctuations in the redshift distribution of galaxies instead of their fluctuation in term of number density, and show how it complements the traditional angular galaxy clustering probe.

All these developments allow to improve our understanding of the matter distribution, from galactic to cosmological scales. These results will allow us to maximize the scientific outputs of next generation surveys and to push our knowledge of cosmological physics to its limits.

My thesis manuscript is organised as follows. I first introduce in Chapter 1 some basic principles of cosmology, from the background and homogeneous evolution of the Universe to the growth of structures. I then present in Chapter 2 my work on the stellar-to-halo mass relation and on the evolution of the cold molecular gas at high redshift. In Chapter 3 I present the cosmological interest of combining *Euclid* main observables with CMB observables. In Chapter 4 I introduce a new cosmological observable and show how it can be used to maximize the outcome of next generation spectroscopic surveys. Finally, Chapter 5 summarizes the results and opens up to future prospects.

# Contents

<b>Remerciements</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>Résumé</b>	<b>v</b>
<b>Synthèse en Français</b>	<b>vii</b>
<b>Introduction</b>	<b>xi</b>
<b>1 From the Big Bang to galaxies</b>	<b>1</b>
1.1 Cosmological background . . . . .	1
1.1.1 Content of the Universe . . . . .	1
1.1.2 The metric of an expanding Universe . . . . .	3
1.1.3 Redshift . . . . .	3
1.1.4 Friedmann equations . . . . .	4
1.1.5 The standard model of cosmology . . . . .	5
1.2 The cosmic microwave background . . . . .	6
1.2.1 The early Universe and the CMB . . . . .	6
1.2.2 Late-time anisotropies . . . . .	6
1.2.3 Summary of CMB observations . . . . .	9
1.3 Summary of upcoming galaxy surveys . . . . .	10
1.4 Growth of inhomogeneities . . . . .	10
1.5 Statistics of the distribution of matter . . . . .	12
1.5.1 Power spectrum . . . . .	12
1.5.2 Angular power spectrum . . . . .	15
1.5.3 Projected observables . . . . .	16
1.6 Non linear scales . . . . .	18
1.6.1 Modelling non linear evolution . . . . .	18
1.6.2 Halo mass function . . . . .	20
1.7 Astrophysical effects on galaxy power spectrum . . . . .	22
1.7.1 Bias of haloes and galaxies . . . . .	22
1.7.2 Redshift space distortions . . . . .	23
<b>2 Galaxy evolution within dark matter haloes</b>	<b>25</b>
2.1 Galaxies and haloes evolution . . . . .	26
2.1.1 Gas cooling . . . . .	26

2.1.2	Galaxy formation efficiency . . . . .	27
2.1.3	Stellar to halo mass relation . . . . .	28
2.2	Galaxies . . . . .	30
2.2.1	The COSMOS field . . . . .	30
2.2.2	Stellar mass function . . . . .	31
2.3	Dark matter haloes . . . . .	32
2.3.1	Halo mass function . . . . .	32
2.3.2	Fit on dark matter simulation . . . . .	33
2.4	Estimating the stellar-to-halo mass relation . . . . .	34
2.4.1	Implementation . . . . .	35
2.4.2	Fitting procedure . . . . .	37
2.4.3	Main sources of SHAM uncertainties . . . . .	38
2.5	Results: stellar-to-halo mass relation in COSMOS . . . . .	38
2.5.1	Stellar-to-halo mass ratio . . . . .	38
2.5.2	Variation of the peak halo mass with redshift . . . . .	39
2.5.3	Impact of the halo mass function on our results . . . . .	41
2.5.4	Stellar to halo mass ratio evolution at fixed halo mass . . . . .	43
2.5.5	Interpretation of this evolution . . . . .	44
2.6	SHMR from a hydrodynamical simulation . . . . .	45
2.7	Cold molecular gas and dark matter haloes . . . . .	46
2.7.1	Linking gas mass to stellar mass . . . . .	46
2.7.2	Linking gas mass to halo mass . . . . .	47
2.7.3	Results . . . . .	48
2.8	What could explain a redshift evolution of quenching ? . . . . .	50
2.9	Conclusions . . . . .	53
<b>3</b>	<b>The cosmological power of a joint analysis of <i>Euclid</i> and CMB surveys</b>	<b>57</b>
3.1	Interests of probe combination . . . . .	58
3.2	Cosmological forecasts . . . . .	60
3.2.1	The likelihood function . . . . .	60
3.2.2	Fisher analysis . . . . .	62
3.2.3	MCMC as an alternative to Fisher . . . . .	63
3.2.4	Fisher matrix using angular power spectra . . . . .	64
3.3	Modelling observables . . . . .	64
3.3.1	Galaxy observables . . . . .	65
3.3.2	CMB lensing . . . . .	66
3.3.3	CMB temperature and polarisation . . . . .	67
3.4	<i>Euclid</i> observables . . . . .	67
3.4.1	Introducing the <i>Euclid</i> mission . . . . .	67
3.4.2	Implementation of the <i>Euclid</i> observables . . . . .	68
3.5	CMB experiments and noise models . . . . .	70
3.5.1	<i>Planck</i> . . . . .	70
3.5.2	Simons Observatory . . . . .	71
3.5.3	CMB-Stage 4 . . . . .	71

3.6	Forecasting cosmological constraints . . . . .	72
3.6.1	Extensions of $\Lambda$ CDM . . . . .	72
3.6.2	Fisher matrix . . . . .	74
3.6.3	Scenarios for the combined analysis . . . . .	75
3.6.4	Limits of the numerical resolution . . . . .	75
3.7	Results . . . . .	77
3.8	Conclusion . . . . .	82
<b>4</b>	<b>An alternative probe for galaxy surveys</b>	<b>85</b>
4.1	Introduction . . . . .	85
4.2	Surveys under consideration . . . . .	86
4.2.1	The DESI experiment . . . . .	86
4.2.2	The <i>Euclid</i> spectroscopic survey . . . . .	86
4.2.3	Tomography . . . . .	87
4.3	Observables . . . . .	88
4.3.1	Galaxy angular density fluctuations . . . . .	88
4.3.2	Angular redshift fluctuations . . . . .	89
4.3.3	Angular power spectra . . . . .	90
4.3.4	Numerical recipes . . . . .	91
4.4	Signal to noise forecasts . . . . .	94
4.5	Fisher forecasts . . . . .	99
4.5.1	Results for the $\Lambda$ CDM model . . . . .	100
4.5.2	Extension to CPL Dark Energy parametrization . . . . .	103
4.5.3	Combining with CMB lensing . . . . .	103
4.6	Discussion . . . . .	107
4.7	Conclusion . . . . .	108
<b>5</b>	<b>Conclusion</b>	<b>111</b>
<b>A</b>	<b>CMB lensing</b>	<b>115</b>
A.1	Introduction . . . . .	115
A.2	Statistical anisotropy . . . . .	116
A.3	Quadratic estimator . . . . .	117
<b>B</b>	<b>Impact of massive neutrinos</b>	<b>119</b>
<b>C</b>	<b>Related publications</b>	<b>123</b>



# Chapter 1

## From the Big Bang to galaxies

### 1.1 Cosmological background

#### 1.1.1 Content of the Universe

Our Universe is made of several components that are interacting together and are affecting its shape and its size. Many independent experiments have made this description of the Universe reliable. However, most of the content of the Universe must be described by components that are not included in our present standard model of physics. As such, cosmology is a huge test and real life laboratory that poses a challenge to our understanding of physics.

The *baryons*<sup>1</sup> are the basic components of matter, as they form the atoms that are in the planets, the stars, the dust and the gas of our Universe. However, baryons account for only five percent of the present day matter-energy<sup>2</sup> budget of our Universe. Another 25% is in the form of *cold dark matter*, where *cold* means that the particles are not relativistic, they move slowly compared to the speed of light. This matter, not included in the standard model of physics, interacts only by gravitation. As such it does not emit light, and is invisible with standard telescopes. Gravitational effects, like dynamics of galaxies or gravitational lensing, allow to detect its presence. The last part of the matter-energy budget, around 70%, is composed of *dark energy*. Dark energy, also unknown to the standard model of physics, is responsible for the acceleration of the expansion of the Universe. It is unclear today if the dark energy takes the form of an unknown fluid, a gravitational constant, or something else.

The last two species of this cosmic inventory are the photons and the neutrinos. Photons mainly come from the relic radiation of the Big Bang. They are travelling freely in the Universe since the moment it became transparent around 380,000 years after the Big Bang. These photons are observed in what we call the cosmic microwave background (CMB). Their energy budget is negligible today, but they used to have a major impact as they were dominating the matter-energy budget of

---

<sup>1</sup>Cosmologists refers all components of atoms as baryons, so both nuclei and electrons. This is technically incorrect (electrons are not baryons), but one can say that the mass of electrons is much lower than the mass of nuclei.

<sup>2</sup>Matter and energy can be compared together thanks to the famous equation  $E = mc^2$ .

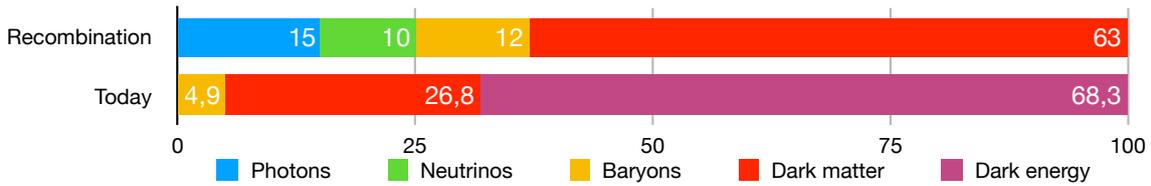


Figure 1.1: Relative importance on the matter-energy budget of the Universe for each of its main component. The upper bar shows the budget at the time of recombination, i.e. the moment of the emission of the CMB, and the bottom bar is the present day budget.

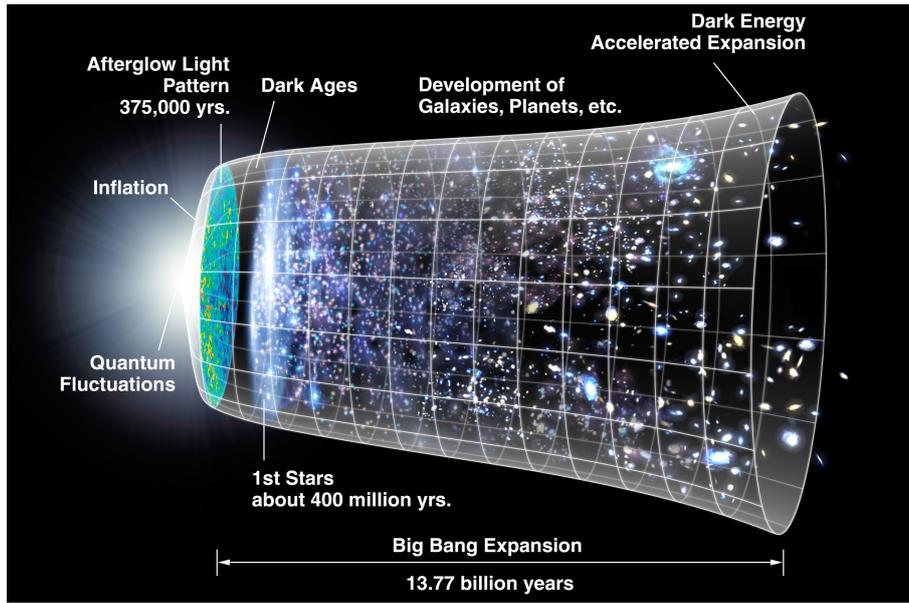


Figure 1.2: Timeline of the Universe. The time goes from left to right, and each slice on this axis is a representation of the Universe at this moment. This Figure represents the different phases of the Universe, from the primordial quantum fluctuations that has been stretched out by inflation, to the present dark energy accelerated expansion. Credits: NASA/WMAP.

the early Universe. Neutrinos are massive particles, even if we only have an upper limit for their mass. They are weakly interacting with matter and their speed is close to the speed of light.

Figure 1.1 shows the relative contribution to the matter-energy budget of the main components of the Universe at two moments in its life: at the time of recombination when the Universe became transparent around 13.5 billion years ago, and today. As we can see, neutrinos and photons disappeared from the matter-energy budget, while dark energy is now dominating.

The Figure 1.2 is a representation of the evolution of the Universe and of the key moments of its existence, from initial quantum fluctuations to present day accelerated expansion. We detail several of these moments in the present chapter to introduce concepts that are of interest for this thesis.

### 1.1.2 The metric of an expanding Universe

The Universe must be described in the framework of general relativity, meaning that we must define a metric. To a certain extent the Universe is homogeneous and isotropic. These hypotheses, called the “cosmological principle”, are validated by observations of the CMB and of the distribution of galaxies at large scales (typically above  $\sim 200 \text{ Mpc}^3$ ). They imply that our position in the Universe is not special and that observers anywhere in the Universe will see the same thing in all directions.

Derived from this homogeneity and isotropy hypothesis, the Friedmann-Lemaître-Robertson-Walker (FLRW) metric is the most general metric for our Universe

$$ds^2 = c^2 dt^2 - a^2(t) d\chi^2, \quad (1.1)$$

where  $t$  is the cosmic time,  $a(t)$  is the scale factor, and  $d\chi^2$  is a comoving distance element on a hypersurface. As our Universe is expanding it is convenient to work in a coordinate system that decouples from the impact of this expansion. These are the comoving coordinates  $\chi$ . The scale factor links the proper (physical) distance  $d$  to the comoving distance by  $d = a(t) \int d\chi$

The curvature of our Universe could be spheric, flat or hyperbolic. A general form of the spatial metric can be expressed in spherical polar coordinates

$$d\chi^2 = \frac{dr^2}{1 - kr^2} + r^2 d\Omega^2, \quad (1.2)$$

where  $d\Omega^2 = d\theta^2 + \sin^2 \theta d\phi^2$ , and  $k = -1, 0$  or  $+1$  if the geometry of the Universe is respectively hyperbolic, flat or spheric. The proper physical distance between us and an object at radial comoving coordinate  $r$  and at a given time  $t$  is then given by

$$d(r, t) = a(t) \int_0^r \frac{dr}{\sqrt{1 - kr^2}}. \quad (1.3)$$

One particular interesting distance is the one travelled by a photon at a time  $t$ . A photon travels along a geodesic following  $ds^2 = 0$ , which gives

$$\chi(t) = \int_0^t \frac{c dt}{a(t)}. \quad (1.4)$$

### 1.1.3 Redshift

Hubble (1929) observed that, in average, galaxies are recessing from us, and the further they are the faster. This led to the discovery of the expansion of the Universe. This expansion is characterized by the Hubble parameter  $H = \dot{a}/a$ . The present value of this parameter  $H_0$  is called the Hubble constant. Due to the expansion, the wavelength of photons emitted by distant galaxies is shifted towards longer wavelengths. This effect is called redshift (noted  $z$ ) and is defined by

$$\frac{\lambda_{\text{obs}}}{\lambda_{\text{emit}}} = 1 + z = \frac{1}{a(t_{\text{emit}})}. \quad (1.5)$$

---

<sup>3</sup>Megaparsecs are a unit of distance:  $1 \text{ Mpc} \simeq 3.086 \times 10^{22} \text{ m}$

Finally, one can rewrite the comoving distance  $\chi$  of Equation 1.4 with

$$\chi(z) = \int_0^z \frac{c dz}{H(z)}. \quad (1.6)$$

To track the evolution of the Universe we then have the choice between different variables that are related, but can all be used as a substitute of the cosmic time. In the future we will use either the redshift, the scale factor or the comoving coordinate  $\chi$ .

### 1.1.4 Friedmann equations

The equation of general relativity of Einstein (1915) links the geometry of space-time (given by the Einstein tensor  $G_{\mu,\nu}$ ) to the matter-energy content of the Universe (included in the stress-impulsion tensor  $T_{\mu,\nu}$ )

$$G_{\mu,\nu} + \Lambda g_{\mu,\nu} = \frac{8\pi G}{c^4} T_{\mu,\nu}, \quad (1.7)$$

where  $G$  is the gravitational constant and  $\Lambda$  is the cosmological constant.

Developing this equation with the FLRW metric and assuming that the Universe is filled by a perfect fluid with a given mass density  $\rho$  and pressure  $p$ , we obtain the Friedmann (1924) equations which govern the expansion of space and the dynamic of the Universe

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G \rho}{3} - \frac{k c^2}{a^2} + \frac{\Lambda c^2}{3}, \quad (1.8)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3}. \quad (1.9)$$

As we saw in Section 1.1.1, our universe is composed of a mixture of fluids. Matter and radiation are assumed to be perfect fluids. Their equation of state is given by  $p = \omega\rho$ , with the equation of state parameter  $\omega = 0$  for matter and  $\omega = 1/3$  for radiation. The cosmological constant  $\Lambda$  can be interpreted as a fluid (which is called dark energy), following an equation of state with parameter  $\omega = -1$  and with a density  $\rho_\Lambda = \Lambda/(8\pi G)$ . This exotic fluid thus has a negative pressure, bringing the counterintuitive idea that the more you compress it, the faster it will contract. The curvature of space can also be interpreted as a fluid with  $\omega = -1/3$  and  $\rho_K = -3k c^2/(8\pi G a^2)$ . Introducing the critical density  $\rho_c = 3H^2/(8\pi G)$ , we define the dimensionless density parameters  $\Omega = \rho/\rho_c$ . The Friedmann equations then give

$$H(z) = H_0 \sqrt{\Omega_k (1+z)^2 + \Omega_m (1+z)^3 + \Omega_\gamma (1+z)^4 + \Omega_\Lambda}, \quad (1.10)$$

where the density parameters  $\Omega$  are taken at their present day value.

Each components of the Universe evolves differently with respect to the scale factor. Figure 1.3 shows the evolution of the density of radiation, matter and dark energy with respect to time. Following this we can determine three different periods

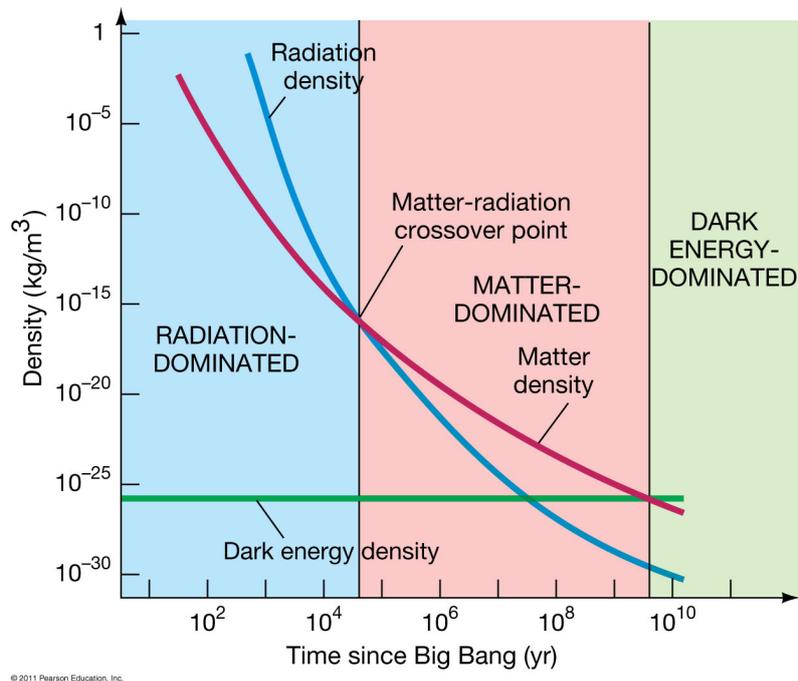


Figure 1.3: Evolution of the densities of radiation, matter and dark energy with time, assuming a flat universe ( $\Omega_k = 0$ ). It is clear here that the Universe went through three different stages: the radiation dominated era, the matter dominated era and finally the dark energy dominated era. Credits: Pearson Education.

on the timeline of the Universe: the radiation dominated era, the matter dominated era, and the present day dark energy dominated era.

### 1.1.5 The standard model of cosmology

The standard model of cosmology is called the  $\Lambda$ CDM model. It assumes that the universe is filled with a dark energy (described by the constant  $\Lambda$ ) and with cold dark matter (CDM). The  $\Lambda$ CDM model also assumes that the space is flat ( $k = 0$ ), and that neutrinos are massless<sup>4</sup>

The flatness of the Universe is supported by strong evidence. Combining observations of the cosmic microwave background with galaxy survey data suggest that  $\Omega_k = 0.001 \pm 0.002$  (Planck Collaboration et al. 2018a). Concerning neutrinos, cosmological analysis give an upper limit on the sum of the mass of neutrinos. The tightest constrains,  $\sum m_\nu < 0.12$  eV, comes from the combination of CMB and galaxy surveys observations (Planck Collaboration et al. 2018a).

The  $\Lambda$ CDM model has been very successful in explaining and predicting a lot of observations, from the CMB power spectrum (Planck Collaboration et al. 2018a) to the Baryonic Acoustic Oscillations (Aubourg et al. 2015). In this thesis we will also introduce extensions of the  $\Lambda$ CDM model, especially in Chapters 3 and 4.

<sup>4</sup>Note that the *Planck* collaboration baseline  $\Lambda$ CDM model assumes massive neutrinos with  $\sum m_\nu = 0.06$  eV, the minimal mass from experiments of neutrino oscillations (Planck Collaboration et al. 2018a). See Appendix B for more details on neutrinos.

## 1.2 The cosmic microwave background

### 1.2.1 The early Universe and the CMB

The most widely accepted theory is that soon after the Big Bang, at a time  $t \lesssim 10^{-35}$  s, the Universe went through an accelerated expansion phase called inflation. Inflation was proposed in the 1980s (Guth 1981; Hawking 1982) to solve at least two major observational problems: the extreme flatness of the Universe, and the homogeneity of the CMB at distances that are not causally connected.

This inflation is also responsible for expanding some tiny quantum fluctuations up to large scales (Mukhanov and Chibisov 1981; Linde 1982). These fluctuations in density and temperature, on top of the homogeneous background, will grow with time to form the anisotropies of the CMB, and later on the large scale structures like galaxies and clusters of galaxies as we know them.

After this inflationary phase and during the radiation dominated era, the Universe continues to expand and to cool down, but at a decelerated rate. Between  $t \sim 3$  and 20 minutes, light atoms like Hydrogen and Helium were created in what is called the Big-Bang nucleosynthesis. At this time, the energy of photons was sufficiently high to prevent electrons and atoms to bind. The Universe was totally opaque, due to the permanent Compton scattering of photons on free electrons. This plasma of baryonic matter and photons was in thermal equilibrium and experienced acoustic oscillations under to the competing effects of gravity and radiative pressure. Meanwhile the dark matter, interacting only by gravity, was not subject to radiative pressure and was thermally decoupled.

At a redshift of  $z \sim 1000$  (around  $t \sim 380,000$  years after the Big Bang) and at a temperature of  $T \sim 3000$  K, photons were cooled enough to allow electrons to recombine with atoms. After this recombination, photons were able to travel freely in the Universe, and they still are today. We are now bathing in a sea of photons coming from this moment of the Universe. This relic radiation is what we call the cosmic microwave background (CMB). The energy spectrum of the photons of the CMB is a perfect black-body, with a present day temperature of  $T_{\text{CMB}} = 2.726$  K. Observations show small anisotropies in the temperature, of the order of  $\Delta T/T = 10^{-5}$  at a scale of 1 degree. Figure 1.4 presents the temperature anisotropies of the CMB as observed by the *Planck* satellite. These anisotropies are the precursor of the present day large scale structures. After recombination the matter is almost unaffected by photons, and under the influence of gravity, baryons will fall into the gravitational potential wells that are formed by the dark matter.

### 1.2.2 Late-time anisotropies

The observed anisotropies of the CMB have multiple origins. They are a combination of *primordial* and *late-time* anisotropies. Primordial anisotropies were seeded by the inflation and evolved under the competing effect of gravity and radiative pressure until the recombination. Late-time anisotropies are created by the interaction between the photons of the CMB and the structures they crossed during their travel. Late-time anisotropies are a secondary effect but they contain a lot of cosmological information. Their analysis shed light on the evolution of the universe

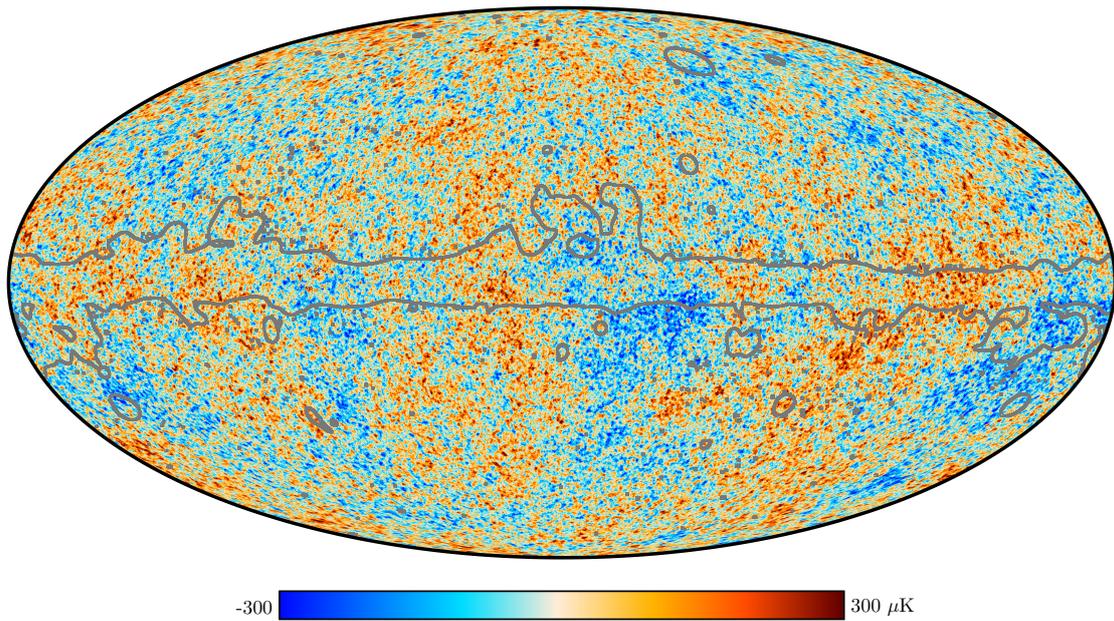


Figure 1.4: The 2018 *Planck* map of the temperature anisotropies of the CMB. The grey outline shows the extent of the confidence mask. Credits: ESA and the *Planck* Collaboration.

from the time of recombination until today. We describe below some of the most important sources of late-time anisotropies, also known as secondary anisotropies (see [Aghanim et al. 2008](#), for a review).

**Reionisation** When the first stars start to shine and to emit photons powerful enough to ionize their environment, the photons from the CMB will be Thomson scattered on these free electrons. The main effects are the suppression of the CMB anisotropy on scales smaller than the horizon size at the epoch of reionisation, corresponding<sup>5</sup> to  $\ell > 20$  ([Bond and Efstathiou 1984](#); [Vittorio and Silk 1984](#)), and the creation of a bump in the polarization power spectrum at large scales, for  $\ell < 20$  ([Liu et al. 2001](#)). The impact of reionisation on the CMB power spectrum is often described with the Thomson optical depth parameter  $\tau$ . Reionisation process may happen inhomogeneously in the Universe, inside growing bubbles centred on the most massive and most star forming galaxies. This patchy reionisation creates secondary anisotropy on small scales (around  $\ell \sim 1000$ ) due to the scattering of CMB photons onto free electrons of ionised gas in galaxies that have a non zero proper velocity compared to the CMB rest frame ([Aghanim et al. 1996](#)).

**Gravitational lensing** The presence of mass along the line of sight will deviate the path of CMB photons due to gravitational effects ([Blanchard and Schneider 1987](#); [Lewis and Challinor 2006](#)). This lensing will distort the primordial image of the

<sup>5</sup>We jump a bit ahead here and introduce the angular scales  $\ell$  which corresponds to multipole moments in the spherical harmonic transformation (a larger  $\ell$  means a smaller angular scale,  $\theta \simeq 180^\circ/\ell$  gives a rough conversion). See Section 1.5.2 for more details.

CMB, and will generate statistical anisotropies. Averaged on the sky, the lensing will smooth the peaks and troughs of the power spectrum, and will create power on the smallest scales ( $\ell > 3000$ ). We introduce in more details the CMB lensing in Chapter 3 and in Appendix A.

**Integrated Sachs-Wolfe effect** This effect is due to the interaction between the CMB photons and a time varying gravitational potential (Sachs and Wolfe 1967). A photon gains energy when falling in a gravitational potential and loses energy when escaping from it. If the size of the potential well decreases while the photons cross it, the energy gained on one side is larger than the one lost on the other side, which causes a net gain of energy. The literature make the distinction between the early integrated Sachs-Wolfe effect (ISW) and the late ISW. The early ISW happened around the time of recombination, when the radiative pressure of photons made the potentials wells decay inside the sound horizon. The late ISW arrived later on, when the Universe expansion accelerated under the domination of dark energy. The ISW effect is mainly seen on the largest scales ( $\ell < 30$ ) because photon crossed very few large wells, while it is negligible of small scales, as this effect cancels out when averaging on a lot of wells and hills.

**Sunyaev-Zeldovich effect** When a photon crosses a dense structure, like a cluster of galaxies, it will scatter on energetic free electrons of the hot gas and gain energy through the Compton-inverse effect (Sunyaev and Zeldovich 1972, 1980). There are two different kinds of Sunyaev-Zeldovich (SZ) effect, the thermal one, when photons scatter on hot electrons and the kinetic one, which is due to the bulk motion of the cloud of electrons. They both have a signature which depends of the frequency, but the kinetic one has the same dependence as the CMB while the thermal one creates spectral distortions which peaks at  $\ell \sim 2000$ .

Moreover, on top of the photons from the CMB there are other sources that emit in the same wavelength range. These will add to the CMB signal and must be removed or properly taken into account before doing any analysis of the primordial or late-time anisotropies. These photons mainly come from:

**The Milky Way** Our galaxy also adds to the extra-galactic emission. There are four main sources of galactic emission: synchrotron emission due to the Bremsstrahlung emission of relativistic electrons spiralling in the galactic magnetic field (Davies and Wilkinson 1998), free-free emission due to free electrons scattering on ions (Bartlett and Amram 1998), thermal emission from cool interstellar dust (Draine 2004), and anomalous microwave emission produced by spinning dust grains (Kogut 1999).

**Other galaxies** They are subject to the same mechanism as our galaxy, but they are often sorted into two categories depending on the dominant effect: radio sources are dominated by synchrotron and free-free emission, they are linked to galaxies with an Active Galactic Nucleus (AGN) (Toffolatti et al. 1999); infrared sources are dominated by the thermal emission from dust, they are dusty star forming galaxies and are responsible for the Cosmic Infrared Background (CIB) (Lagache et al. 2005).

The observed image of the CMB is a combination of all of these effects, and we need to decouple and isolate them to be able to measure the primordial CMB. These effects have both an impact on the temperature of the photons and on their polarization.

The late-time anisotropies can be seen as a contamination of the primordial image of the CMB, but they are also a huge source of information on the structures and content of the Universe through all the stages of its evolution since the recombination. These late-time anisotropies can be used as powerful cosmological probes that provide additional information on the large scale structures. Two examples among many others: the reconstructed lensing signal (Planck Collaboration et al. 2018b), or cluster counts detected by SZ effect (Planck Collaboration et al. 2016b).

### 1.2.3 Summary of CMB observations

Since its discovery by Penzias and Wilson (1965), the CMB has been extensively observed, especially by satellites like the Cosmic Background Explorer (COBE, Smoot et al. 1992), the *Wilkinson* Microwave Anisotropy Probe (WMAP, Larson et al. 2011) and *Planck* (Planck Collaboration et al. 2018c), the most recent and precise to date.

These satellite observations are completed by ground-based telescopes such as the South Pole Telescope (SPT, Story et al. 2013) and the Atacama Cosmology Telescope (ACT, Sievers et al. 2013). They usually have a higher resolution than space based telescope: for example ACT has a beam size of roughly 1.4 arcmin, versus 7 arcmin for *Planck*. However they observed a smaller patch of the sky: ACT covered  $600 \text{ deg}^2$  and SPT covered  $2500 \text{ deg}^2$  while *Planck* observed the full sky, i.e.  $\sim 41\,000 \text{ deg}^2$ .

Upgraded versions of the instruments mounted on these telescopes, ACTPol (Louis et al. 2017) and SPTPol (Keisler et al. 2015), focused particularly on measuring the polarization of the CMB, together with other dedicated telescopes such as Polarbear (Polarbear Collaboration et al. 2014) and BICEP2 (BICEP2 Collaboration et al. 2014). The targets are the primordial B modes of polarization, with the scope of observing primordial gravitational waves and constrain scenarios of inflation.

New generation experiments are currently ongoing, such as Advanced ACT (Henderson et al. 2016; Aiola et al. 2020), SPT-3G (Benson et al. 2014), the Simons Array (follow up of Polarbear, Suzuki et al. 2016), BICEP3 (Ahmed et al. 2014), or the Cosmology Large Angular Scale Surveyor (CLASS, Harrington et al. 2016). They all should provide results in the coming years.

The future of CMB observations is already full of promises. The Simons Observatory (The Simons Observatory Collaboration et al. 2019) a next generation ground based telescope, is currently being built and should began operations in the early 2020s. Litebird (Matsumura et al. 2014) is a satellite currently into development, targeting the detection of primordial gravitational waves by performing a full-sky survey of the polarization of the CMB. Plans for a more distant future are already laid down with the CMB Stage 4 telescope (Abazajian et al. 2019), which should starts operations in the late 2020s.

### 1.3 Summary of upcoming galaxy surveys

In the coming years, large scale optical and infra-red (IR) surveys will map with unprecedented accuracy our Universe from the present epoch up to when it was roughly one tenth of its current age. A significant part of these surveys will be spectroscopic, e.g., DESI (DESI Collaboration et al. 2016), 4MOST (de Jong and 4MOST Consortium 2015), WEAVE (Bonifacio et al. 2016), or *Euclid* with the NISP instrument (Laureijs et al. 2011) and will provide us with spectra for large samples of sources. Such spectra will not only enable deep insight on the physics of those objects, but will also yield accurate estimates of their redshift and thus of their distance to the observer. From the cosmological point of view, this will enable a precise (statistical) characterization of the (apparent) spatial distribution of those luminous tracers (via two- or three-point statistics), and this itself should shed precious light on open topics such as, e.g., the nature of dark energy, the possible interplay of dark energy and dark matter, the mass hierarchy of neutrinos, or possible deviations of gravity from General Relativity, to quote a few.

At the same time, a different family of surveys will scan the sky at greater depths with optical filters, and with exquisite image quality. These photometric experiments build high quality, very large catalogues of sources, with however relatively rough redshift estimations given their moderate number of filters/colours. While mining the faint Universe, this type of surveys will be particularly sensitive, from a cosmological perspective, to the angular clustering of luminous matter, the cosmological aspects of gravitational lensing throughout cosmic epochs, the satellite population in haloes, and the formation and evolution of the population of galaxy clusters. In this context, the Dark Energy Survey (DES, Abbott et al. 2018) is currently providing state-of-the-art cosmological constraints in the late universe, and these should be further complemented by the Vera Rubin Observatory (LSST, Ivezić et al. 2019), which, at the same time, will also explore the variability of the night sky in a regime of depth and time domain that so far remains practically unexplored.

An intermediate third class of experiments also exists. These are the *spectro-photometric* surveys that conduct standard photometry in a relatively large set (from  $\sim 10$  up to  $\sim 60$ ) of narrow-band optical filters. This strategy combines the indiscriminate character of the photometric surveys with *high* precision redshift estimates ( $\Delta z/(1+z) \sim 10^{-3}$ – $10^{-2}$ ) for a large fraction ( $> 20$ – $30$  %) of the detected sources. Given its multi-color character, these surveys are able to provide *pseudo-spectrum* / *photo-spectrum* in each pixel of the surveyed area. The pioneer example of COMBO-17 has been or is being followed by other efforts such as COSMOS (Scoville et al. 2007a), ALHAMBRA (Moles et al. 2008), SHARDS (Pérez-González et al. 2013), PAU (Martí et al. 2014), J-PAS (Benitez et al. 2014), SPHEREx (Doré et al. 2014), and J-PLUS (Cenarro et al. 2019).

### 1.4 Growth of inhomogeneities

Even if the Universe on large scales is well described by the Friedmann equations derived in Section 1.1.4, we saw that the CMB is anisotropic, and observation of galaxies showed that they are not homogeneously distributed on intermediate scales

(e.g. [de Lapparent et al. 1986](#)). We derive in this section the general formalism to describe the growth of structures from the small initial perturbation that are present after the inflation up to the large scale structures as we know them.

We start from a simple case: the Newtonian and linearised case. The Newtonian regime simplifies the equations of general relativity we saw in Section 1.1.4, and is valid as long as we consider a region smaller than the Hubble horizon  $cH^{-1}$ . For a more detailed description of the hypotheses and derivations see e.g. [Peebles \(1980\)](#).

We define the comoving coordinates  $\mathbf{x}$  of a position in space such that the physical coordinate is  $\mathbf{r} = \mathbf{x}a$ . The proper velocity relative to the origin is  $\mathbf{u} = \dot{a}\mathbf{x} + \mathbf{v}(\mathbf{x}, t)$ , and  $\mathbf{v}$  is the peculiar velocity. We define the overdensity field  $\delta$  as the fluctuation around the mean density

$$\rho(\mathbf{x}, t) = \bar{\rho}(t) (1 + \delta(\mathbf{x}, t)) \quad (1.11)$$

We consider matter as an ideal fluid with a potential  $\phi$ . The general fluid equations in comoving coordinates are given by

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{1}{a}(\mathbf{v} \cdot \nabla)\mathbf{v} + \frac{\dot{a}}{a}\mathbf{v} = -\frac{1}{\rho a}\nabla p - \frac{1}{a}\nabla\phi, \quad (1.12)$$

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla \cdot (1 + \delta)\mathbf{v} = 0, \quad (1.13)$$

$$\frac{1}{a^2}\nabla^2\phi = 4\pi G \bar{\rho} \delta. \quad (1.14)$$

We assume small perturbations in the density ( $\delta \ll 1$ ) and small perturbation in the peculiar velocity. This assumes that  $(\frac{vt}{d})^2 \ll \delta$ , with  $d$  the coherence length for the spatial variation of  $\delta$ , and  $t \sim (G\bar{\rho})^{-1/2}$  the typical variation time of an overdensity.

Developing the perturbation equations above to first order and combining them, we obtain

$$\frac{\partial^2 \delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial \delta}{\partial t} = \frac{\nabla^2 p}{\bar{\rho} a^2} + 4\pi G \bar{\rho} \delta, \quad (1.15)$$

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla \cdot \mathbf{v} = 0. \quad (1.16)$$

We see here that the density will oscillate under the combination of two forces, pressure and gravity.

Let us now make the hypothesis of an adiabatic evolution of the fluid, i.e. no heat is exchanged between particle elements. Introducing the speed of sound  $c_s = \sqrt{\partial P / \partial \rho}$ , adiabatic evolution brings the following relation between pressure and density

$$\frac{1}{\rho}\nabla P = \frac{1}{\rho}c_s^2\nabla\rho = \frac{c_s^2}{1+\delta}\nabla\delta. \quad (1.17)$$

Equation 1.15 now becomes

$$\frac{\partial^2 \delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial \delta}{\partial t} = 4\pi G \bar{\rho} \delta + \frac{1}{a^2}c_s^2\nabla^2\delta. \quad (1.18)$$

We can develop the density field into plane waves  $\delta(\mathbf{x}, t) = \sum \delta_k(t) \exp i \mathbf{x} \cdot \mathbf{k}$ , and  $\lambda = 2\pi a/k$  is the physical wavelength. This development put in Equation 1.18 brings the Jeans length defined by

$$\lambda_J = c_s \sqrt{\frac{\pi}{G \bar{\rho}}} . \quad (1.19)$$

All scales above the Jeans length will collapse, as the gravity dominates over pressure. For smaller scales, the attraction of gravity is countered by the pressure of the fluid, and the overdensity will oscillate. During the radiation domination era, it is the radiative pressure of photons that prevents the collapse of the small scales. This phenomenon is responsible for the Baryonic Acoustic Oscillation (BAO), a typical scale at which the matter is distributed, and which is still present in the distribution of galaxies, long after the decoupling between matter and radiation (Eisenstein et al. 2005).

During the matter domination era (and in the scenario of a weakly interacting dark matter) the pressure can be neglected. This means that all scales will collapse as they are all above the Jean's length. Neglecting the pressure term, Equation 1.15 can be written

$$\frac{\partial^2 \delta}{\partial t^2} + 2 \frac{\dot{a}}{a} \frac{\partial \delta}{\partial t} - 4\pi G \bar{\rho} \delta = 0 , \quad (1.20)$$

which is the equation of an harmonic oscillator without any spatial derivatives anymore. It is then possible to decouple the density evolution into a spatial and a temporal evolution,  $\delta(\mathbf{x}, t) = D(t)\delta_0(\mathbf{x})$ . The time component  $D$  is called the growth factor and follows the equation

$$\frac{d^2 D}{dt^2} + 2 H \frac{dD}{dt} - \frac{3}{2} H^2 \Omega D = 0 . \quad (1.21)$$

The solution of the above equation is the sum of a growing and a decaying mode, so we have

$$\delta(\mathbf{x}, t) = D^+(t) \delta^+(\mathbf{x}, 0) + D^-(t) \delta^-(\mathbf{x}, 0) . \quad (1.22)$$

For an Einstein-de Sitter Universe, i.e. during the matter domination, an exact solution is given by  $D^+(t) \propto t^{3/2}$  and  $D^-(t) \propto t^{-1}$ .

As we see from Equation 1.20 and 1.22, it is possible to decouple the spatial and the time evolution of an overdensity. This leads to a powerful way of measuring the evolution of the Universe. As long as we are in the linear regime and for scales that are below the horizon size, the distribution of matter at any time can be computed by multiplying the initial distribution of matter by the growth factor.

## 1.5 Statistics of the distribution of matter

### 1.5.1 Power spectrum

From the Friedmann equations (section 1.1.4), we were able to model the growth of matter perturbations (section 1.4). We now need a tool to test our model against observations. This tool is the correlation function. The main advantage of the correlation function (or its Fourier transform, the power spectrum), lies in the fact that

it allows to compress information of the matter distribution in only one statistical tool. Moreover as long as we are interested in large scales this compression does not lose information, because these scales are in the linear regime. In this regime the overdensity field is Gaussian and isotropic, and is fully described by its two-point correlation function.

In this section we derive relations for the matter (including both dark and baryonic) distribution. These hold also for e.g. the distribution of galaxies. We will later assume that galaxies are sampling the underlying distribution of matter, with some bias. We will discuss about this and the corresponding galaxy bias in Section 1.7.1.

We are interested in the over-density field  $\delta(\mathbf{x}, t)$ , as defined in Equation 1.11. We introduce the *two-point correlation function* as the covariance of  $\delta(\mathbf{x}, t)$  between two positions in the universe with a comoving separation  $\mathbf{r}$ ,

$$\xi(\mathbf{r}, t) = \langle \delta(\mathbf{x} + \mathbf{r}, t) \delta(\mathbf{x}, t) \rangle_{\mathbf{x}} , \quad (1.23)$$

where  $\langle \rangle_{\mathbf{x}}$  denotes an average over the comoving coordinates  $\mathbf{x}$ . As the Universe is statistically homogeneous and isotropic, the correlation function depends only on  $r$ , the modulus of  $\mathbf{r}$ .

The Fourier transform of the correlation function is called the *power spectrum*, and is given by

$$P(k, t) = \int d^3\mathbf{r} \xi(r, t) \exp(i\mathbf{k} \cdot \mathbf{r}) . \quad (1.24)$$

Similarly, the power spectrum can be obtained as the covariance between two modes of the Fourier transform of the density field noted with a tilde  $\tilde{\delta}$ . As the Universe is statistically homogeneous and isotropic, we get

$$\langle \tilde{\delta}(\mathbf{k}, t) \tilde{\delta}(\mathbf{k}', t) \rangle = (2\pi)^3 P(k, t) \delta_{\text{D}}(\mathbf{k} - \mathbf{k}') , \quad (1.25)$$

where  $\delta_{\text{D}}$  is the Dirac delta function.

In the linear regime, each Fourier mode evolves independently, because Equation 1.21 does not involve spatial derivatives. The linear power spectrum  $P(k, t)$  can be obtained from the linear growth of the power spectrum at the end of recombination.

The initial power spectrum of the gravitational potential at the end of inflation is assumed to be scale invariant. Linking the potential to the density with the Poisson equation gives

$$P_{\text{init}}(k) = A_s k^{n_s} , \quad (1.26)$$

with  $A_s$  a constant and  $n_s \approx 1$  the scalar spectral index.

Inflation creates fluctuations with size larger than the horizon. These fluctuations grow through self-gravity. As time goes, the size of the horizon grows, and new modes are entering the horizon. As long as radiation dominates over gravity, these modes are frozen: the universe is expanding too rapidly under the effect of radiation for the matter fluctuations to collapse under gravity. When matter and radiation density reach equality, dark matter fluctuations can grow exponentially. Baryonic matter and photons start oscillating under the competing effect of gravity and pressure. After recombination, pressure is negligible and all modes start to grow. The evolution of the primordial power spectrum during the radiation dominated era is

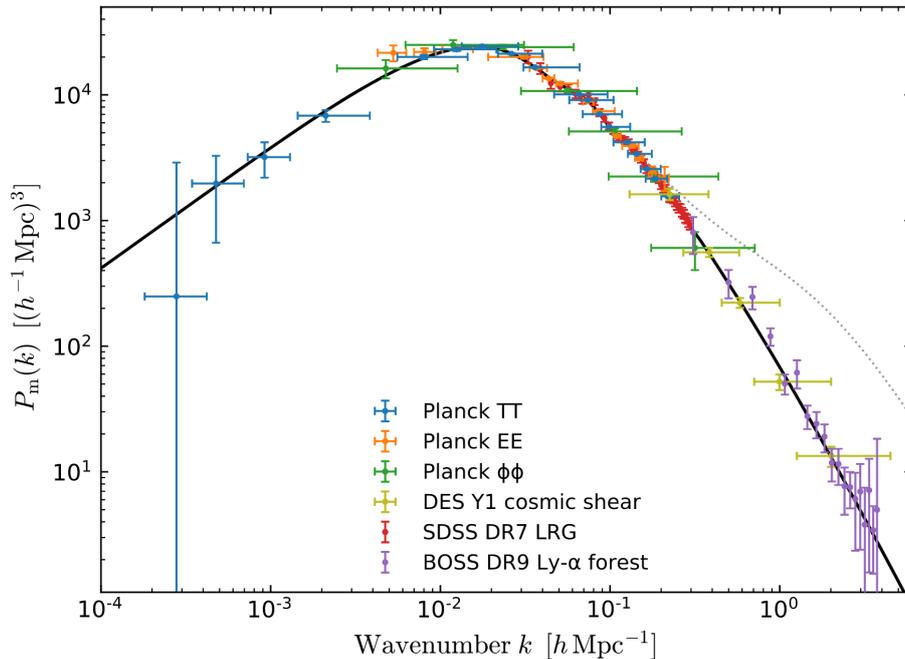


Figure 1.5: Figure from [Planck Collaboration et al. \(2018c\)](#). Inference of the linear matter power spectrum at  $z = 0$  from different measurements. The plain black line is the best fit power spectrum from *Planck* measurements. The agreement on such a large range of scales and redshifts is a huge credit to the predicting power of the  $\Lambda$ CDM model. The dotted line shows the non linear power spectrum for reference.

encoded in the transfer function  $T(k)$ . The power spectrum after recombination, at the beginning of the matter domination era, is given by

$$P_0(k) = A_s k^{n_s} T^2(k) . \quad (1.27)$$

An accurate description of the transfer function can be found in [Bardeen et al. \(1986\)](#).

After recombination, matter decouples from radiation and at linear scales all modes evolve independently and grow following Equation 1.22. The linear power spectrum is then

$$P(k, t) = A_s D_+(t)^2 T^2(k) k^{n_s} . \quad (1.28)$$

Figure 1.5 shows the linear power spectrum estimated by *Planck*. On large scales the power spectrum is proportional to  $k^{n_s}$ , following the power spectrum at the end of inflation, whereas on small scales it evolves like  $k^{-3}$ . The inflexion point corresponds to the size of the Hubble horizon at time of the matter-radiation equivalence.

As we can see in Figure 1.5, for small scales ( $k \gtrsim 0.2 h \text{ Mpc}^{-1}$ ) Equation 1.28 does not hold anymore as the growth of structures is non linear. We discuss non linear evolution in Section 1.6.

### 1.5.2 Angular power spectrum

The CMB is observed as a 2D field. We can also consider galaxy surveys as 2D fields if we bin them by intervals of redshift. The 2-point correlation function on the sphere is called the angular correlation function  $w(\theta)$ , which is the analogous of the correlation function  $\xi(r)$  of Equation 1.23. The analogous of the power spectrum  $P(k)$  on the sphere is called the angular power spectrum and is noted  $C_\ell$ . It is obtained by decomposing the observed field onto the spherical harmonics basis.

The spherical harmonics  $Y_{\ell m}(\theta, \phi)$  are the solution of the Laplace equation in spherical coordinates,  $\Delta Y_{\ell m} = 0$ . These functions are orthogonal, with normalization

$$\int d\Omega Y_{\ell m}^*(\theta, \phi) Y_{\ell' m'}(\theta, \phi) = \delta_D(\ell - \ell') \delta_D(m - m'). \quad (1.29)$$

where  $d\Omega = \sin\theta d\theta d\phi$ . See e.g. Dodelson (2003) for more details.

We can decompose any spherical 2D field  $f(\theta, \phi)$  (such as the observed CMB temperature) on this basis

$$f(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\theta, \phi), \quad (1.30)$$

and the coefficients  $a_{\ell m}$  are given by

$$a_{\ell m} = \int_{\Omega} d\Omega f(\theta, \phi) Y_{\ell m}^*(\theta, \phi). \quad (1.31)$$

The information contained in the field  $f$  is equal to the one contained in its  $a_{\ell m}$ . As such, if the field is Gaussian and has zero mean, like the CMB temperature fluctuations, all the information about this field is contained in the variance of the  $a_{\ell m}$ , which are called  $C_\ell$

$$\langle a_{\ell m} a_{\ell' m'}^* \rangle = \delta_D(\ell - \ell') \delta_D(m - m') C_\ell. \quad (1.32)$$

The CMB we observe is a given random realization of all possibilities. For a given  $\ell$ , the observed  $a_{\ell m}$  are drawn from a probability distribution function which is Gaussian, with mean zero and with a variance  $C_\ell$ . To recover the true underlying  $C_\ell$  from the observed field we can define an estimator  $\hat{C}_\ell$  by

$$\hat{C}_\ell = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} a_{\ell, m} a_{\ell, m}^*. \quad (1.33)$$

Our estimator is limited by the number of multipoles  $m$  for each  $\ell$ . As such there is a fundamental uncertainty in the estimation of the  $C_\ell$  which is called the *cosmic variance*, and is quantified by

$$\left( \frac{\Delta C_\ell}{C_\ell} \right)_{\text{c.v.}} = \sqrt{\frac{2}{2\ell + 1}}. \quad (1.34)$$

Figure 1.6 presents the CMB angular power spectrum of the temperature field

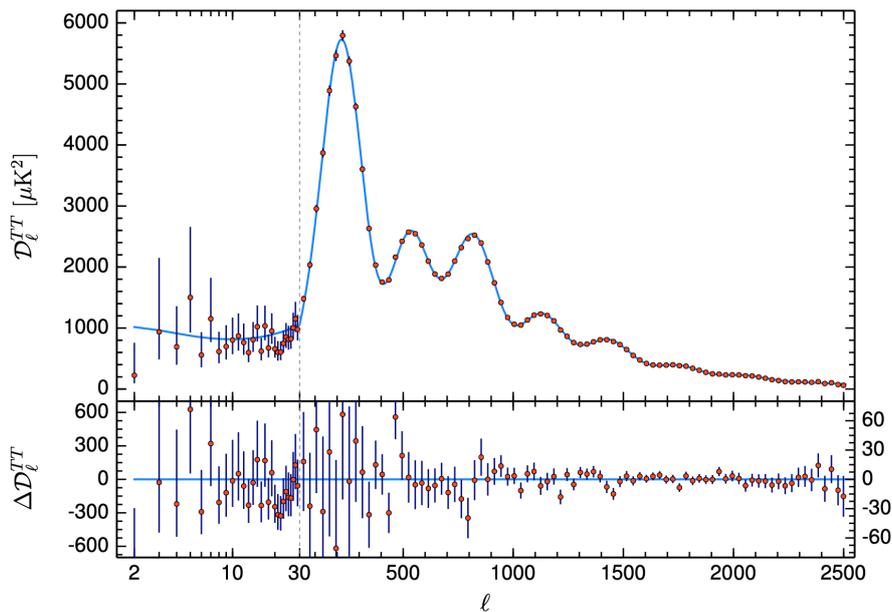


Figure 1.6: Figure from [Planck Collaboration et al. \(2018a\)](#). Temperature power spectrum of the CMB measured by *Planck*, with  $D_\ell \equiv \ell(\ell + 1)C_\ell/(2\pi)$ . The  $\Lambda$ CDM theoretical spectrum best fit is plotted in light blue in the upper panel. Residuals with respect to this model are shown in the lower panel.

measured by the *Planck* experiment ([Planck Collaboration et al. 2018a](#)). This shows again the great agreement of the  $\Lambda$ CDM model with observations.

### 1.5.3 Projected observables

For galaxy surveys, especially when we do not have enough resolution on the redshift measurements, it is common to study the distribution projected on a 2D sphere. We then need to link the observable which is the angular power spectrum  $C_\ell$  to the power spectrum  $P(k)$  obtained from the cosmological model. If we project the 3D matter distribution onto a 2D sphere, under a selection function  $\phi(\chi)$  normalized to unity, we obtain the 2D overdensity field

$$\delta^{2D}(\mathbf{n}) = \int_0^{\chi_\infty} d\chi \phi(\chi) \delta(\mathbf{x}), \quad (1.35)$$

where  $\chi$  is the comoving distance (Equation 1.4) and  $\mathbf{n}$  is unit vector defining a direction on the sky, and  $\mathbf{x} = \chi\mathbf{n}$  is the galaxy position. We integrate up to the maximum comoving distance covered by the survey  $\chi_\infty$ . The 2D overdensity field is noted  $\delta^{2D}$  while its 3D version is noted  $\delta$ . Note that in practice for galaxy surveys there is one more step to link the galaxy density field to the matter density, as galaxies are a biased tracer of matter. We describe this step in more detail in Section 1.7.1.

We develop  $\delta$  on its Fourier coefficients

$$\delta(\mathbf{x}) = \int_{\mathbb{R}^3} \frac{d\mathbf{k}}{(2\pi)^3} \exp(-i\mathbf{k} \cdot \mathbf{x}) \tilde{\delta}(\mathbf{k}), \quad (1.36)$$

where a tilde denotes the Fourier transform.

Using the Equation 1.25, the 2 point correlation function is given by

$$\begin{aligned} \langle \delta^{2D}(\mathbf{n}) \delta^{2D}(\mathbf{n}') \rangle &= \int_0^{\chi_\infty} d\chi \int_0^{\chi_\infty} d\chi' \phi(\chi) \phi(\chi') \\ &\times \int_{\mathbb{R}^3} \frac{d\mathbf{k}}{(2\pi)^3} \exp(i\mathbf{k} \cdot \mathbf{x}) \exp(-i\mathbf{k} \cdot \mathbf{x}') P(k; \chi, \chi'), \end{aligned} \quad (1.37)$$

where  $P(k; \chi, \chi')$  denotes the power spectrum between two different epochs. In the linear theory, it is easy to decouple this power spectrum into  $P(k; \chi, \chi') = P(k) D(\chi) D(\chi')$  where  $D$  is the growth factor. Note that here we take the growth as a function of the comoving distance, whereas it was first described as a function of time in Equation 1.22. This is not a problem thanks to the bijection between time and comoving distance (Equation 1.4).

We then make use of the plane wave expansion, introducing the Legendre polynomials  $P_\ell$  and the spherical Bessel functions  $j_\ell$

$$\exp(i\mathbf{k} \cdot \mathbf{x}) = \sum_{\ell=0}^{\infty} i^\ell (2\ell + 1) P_\ell(\mathbf{k} \cdot \mathbf{x}) j_\ell(k \chi) \quad (1.38)$$

$$= 4\pi \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} i^\ell Y_{\ell m}(\mathbf{k}/k) Y_{\ell m}^*(\mathbf{n}) j_\ell(k \chi). \quad (1.39)$$

Using the above relations and the properties of the spherical harmonic transform defined in Section 1.5.2, we obtain the expression of the angular power spectrum

$$C_\ell = \frac{2}{\pi} \int dk k^2 P(k) \left[ \int_0^{\chi_\infty} d\chi \phi(\chi) D(\chi) j_\ell(k \chi) \right]^2 \quad (1.40)$$

This relation links our observable, the angular power spectrum, to the predictions of the cosmological model made on the power spectrum and on the growth function.

One common approximation in the computation of the angular power spectrum is the Limber (1953) approximation. For high  $\ell$  (typically  $\ell > 100$ ) the power spectrum varies slowly compared to the spherical Bessel function. In this sense the Bessel will extract the typical scale  $k = \frac{\ell+1/2}{\chi}$  (see e.g. Loverde and Afshordi 2008). Using

$$\int dk k^2 j_\ell(k \chi) j_\ell(k \chi') = \frac{\pi}{2\chi^2} \delta_D(\chi - \chi'), \quad (1.41)$$

we can remove some of the integrals of Equation 1.40 and we obtain

$$C_\ell = \int_0^{\chi_\infty} \frac{d\chi}{\chi^2} [\phi(\chi) D(\chi)]^2 P\left(k = \frac{\ell+1/2}{\chi}\right). \quad (1.42)$$

The main interest of this approximation is to reduce the computational time, because due to the oscillatory nature of Bessel functions, accurate computation of angular power spectra is time-consuming.

## 1.6 Non linear scales

### 1.6.1 Modelling non linear evolution

At small scales ( $k > 0.1 h \text{ Mpc}^{-1}$ ), and when the overdensity  $\delta$  becomes of the order of unity, the linearisation of the equations performed in Section 1.4 is not valid anymore. In order to model the non-linear evolution, different approaches have been developed.

Perturbation theory is an analytical development of the perturbed equations to higher orders. Every higher order development is increasingly complex, and in practice it is difficult to go beyond order three for two point statistics (Bernardeau et al. 2002). Typically, the maximum scale reached with precision by the perturbation theory at second order is of  $k_{\text{max}} = 0.2 h \text{ Mpc}^{-1}$  (Taruya et al. 2012). These developments allow to obtain an analytic evolution of the density field for the weakly non-linear regime.

Another successful approach is to rely on dark matter simulations. These numerical simulations (also called N-body simulations) reproduce the distribution of matter by tracking the evolution of dark matter particles. They are mainly limited by the computational power and the trade-off between the resolution and the size of the simulation box. The improvement of dark matter simulation is tightly linked with the development of high performance computing. Simulations like Millennium (Springel 2005), HORIZON 4 $\pi$  (Teyssier et al. 2009), and more recently the Flagship simulation (Potter et al. 2017) allowed to measure the power spectrum from large to small scales, with increasing precision. For instance, the Flagship simulation includes  $2 \times 10^{12}$  particles inside a box with a comoving side length of  $3780 h^{-1} \text{ Mpc}$ . However, as shown in Figure 1.7, when comparing the outcome of different numerical simulations we see that the power spectrum estimation agree at the order of 1% only for scales larger (smaller  $k$ 's) than  $k_{\text{max}} \sim 1 h \text{ Mpc}^{-1}$  at  $z \lesssim 1$ . Figure 1.8 compares the maximum scales where perturbation theory and numerical simulations are considered as reliable.

A third approach is the halo model of large scale structures (Cooray and Sheth 2002). This development assumes that all the matter is contained inside virialised spheres called haloes. This idea is based on the spherical collapse model, which we introduce below in Section 1.6.2. Briefly, the main hypothesis of the halo model is to separate the two-point correlation function into a one-halo term and a two-halo term. The first corresponds to correlations inside one halo, and describes the small scales ( $k > 1 h \text{ Mpc}^{-1}$ ), and the second corresponds to correlations between two haloes and describes the large scales ( $k < 0.1 h \text{ Mpc}^{-1}$ ). The ingredients of the halo model are the abundance and the spatial distribution of haloes and their radial density profile. The key advantage of the halo model is its flexibility. Its framework can be applied to a variety of observables such as galaxy clustering (e.g. Lacasa and Rosenfeld 2016) or tSZ analysis (e.g. Salvati et al. 2018) to cite only two.

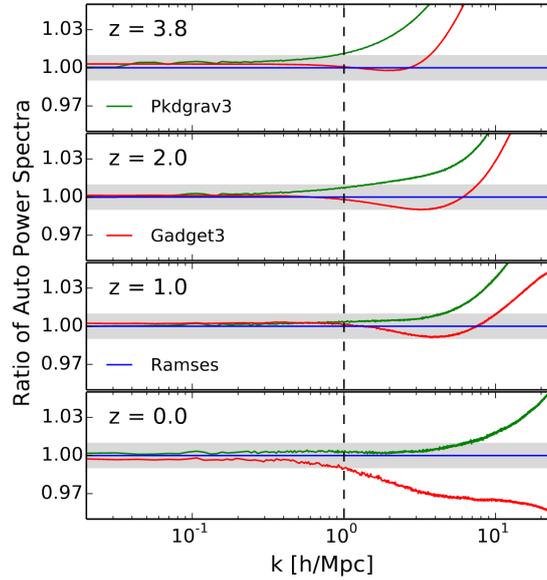


Figure 1.7: Figure from Schneider et al. (2016). Comparison of auto power spectra from three different N-body codes at different redshifts. Green lines correspond to Pkdgrav3 (Potter et al. 2017), red lines to Gadget3 (Springel 2005), and blue lines to Ramses (Teyssier 2002)(reference lines). One percent agreement (indicated by the grey band) is obtained for  $k \lesssim 1 h \text{ Mpc}^{-1}$  (dashed vertical line).

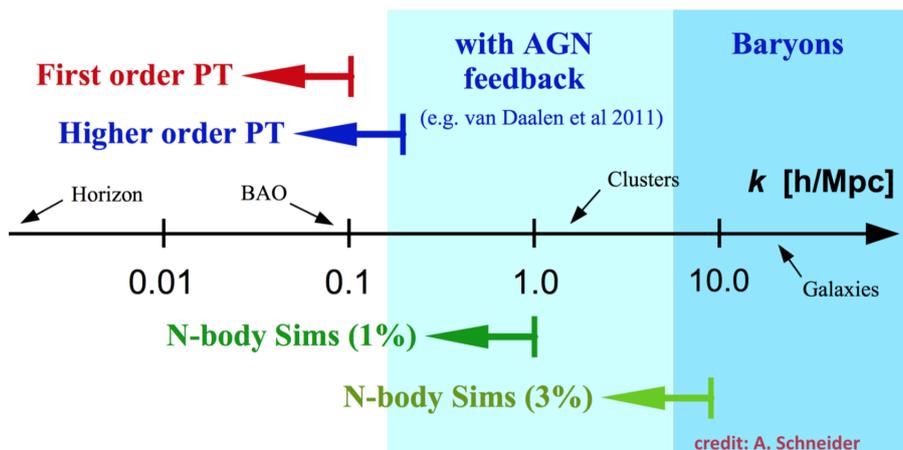


Figure 1.8: Summary of the non linear scales reached by perturbation theory (PT) and numerical N-body simulations. Credits to A. Schneider and S. Codis.

One just has to implement the density profile, abundance, and clustering properties corresponding to the observable of interest.- One drawback of this approach is that it must rely on external data (simulations or observations) to calibrate the halo mass function (see Section 1.6.2), one of the main ingredients of this model.

### 1.6.2 Halo mass function

A first step to model the non linear regime is to study the collapse of a spherically symmetric density distribution. In this configuration, an exact solution of the evolution equations can be obtained, assuming an Einstein-de Sitter universe (Gunn and Gott 1972). Tracking the evolution of the density inside a collapsing sphere with a constant mass, one can show that the sphere collapses (the radius tends to zero) when the linear overdensity reaches  $\delta_c = 1.686$ . Even if not mathematically correct (it is an extrapolation of the linear equations), the  $\delta_c$  value is used as a criterion to describe collapsed objects: each time an overdensity reaches this threshold, it can be considered as collapsed, and the linear equations does not hold anymore.

In practice the sphere will not collapse to a point, but the particles will relax and reach the *virial equilibrium* when the total energy equals half of the potential energy. The ratio between the density inside a virialized sphere and the mean density of the universe (for an Einstein-de Sitter universe, i.e.  $\Omega_m = 1$ ) is equal to (White 2001)

$$\Delta_V \equiv \frac{\rho(r < r_{\text{vir}})}{\bar{\rho}} = 18\pi^2 \simeq 178. \quad (1.43)$$

This development of the non linear evolution of the spherical model led to the idea that we can describe dark matter haloes as virialised spherical objects characterised only by their mass. The statistics that encompass this information is the *halo mass function* (HMF), which gives the number density of virialised haloes as a function of their mass.

The Press-Schechter formalism (Press and Schechter 1974) combines the linear evolution of the density field with the overdensity criterion for the spherical collapse as defined above. The basic idea is to assume that the initial overdensity field is a Gaussian random field, which evolves linearly. Each time a perturbation reaches the critical overdensity  $\delta_c = 1.686$ , it collapses into a virialised halo. This leads to the following expression of the halo mass function

$$\frac{d^2n}{dM dV}(M, z) = \sqrt{\frac{2}{\pi}} \frac{\delta_c \bar{\rho}}{M} \left| \frac{d \ln \sigma_M}{dM} \right| \exp\left(-\frac{\delta_c^2}{2\sigma_M^2}\right). \quad (1.44)$$

This function gives the number of dark matter haloes per unit of mass and volume for this density threshold. The parameter  $\sigma_M$  corresponds to the root mean square (rms) of the fluctuation of the density field smoothed at a scale  $R(M) = (3M / (4\pi \bar{\rho}))^{1/3}$ .

This mass function is shown in Figure 1.9. We see that the HMF scales as  $M^{-2}$  for low masses, and presents an exponential cutoff at large masses, i.e. for  $M > M_*$  where  $M_*$  is given by  $\sigma_M(M_*) = \delta_c$ . This behaviour is related to the hierarchical structure formation scenario: low mass haloes form first, and high mass haloes appear later by merging of smaller haloes or by accretion of matter (Lacey

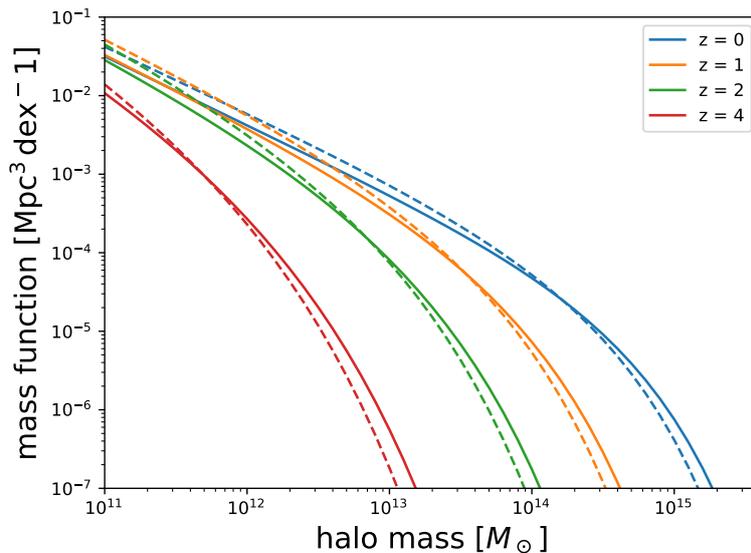


Figure 1.9: Halo mass functions at different redshifts. The plain lines are the [Tinker et al. \(2008\)](#) HMF, while the dashed line are the [Press and Schechter \(1974\)](#) HMF.

[and Cole 1993, 1994](#)). The most massive haloes are very rare structures that are found only in the very dense environments. Because of the dependence of the HMF to initial conditions of the density field and to the growth of structures, observations of the halo mass function are a powerful cosmological probe ([Press and Schechter 1974](#); [Bond et al. 1991](#); [Lucchin and Matarrese 1988](#)).

Since [Press and Schechter \(1974\)](#), improvements have been made in the functional form of the halo mass function. In their work, they used 1000 particles of an N-body simulation to validate their results, while [Sheth and Tormen \(1999\)](#) used  $10^7$  particles and replaced the spherical collapse scenario with an ellipsoidal collapse. [Tinker et al. \(2008\)](#) uses 22 independent N-body simulations with  $\sim 10^9$  particles to obtain a fit of functional form of the HMF. More recently ([Despali et al. 2016](#)) used 6 N-body simulations to test the universality (i.e. its robustness to the halo definition) of the halo mass function. The use of different simulations, from small volume with high resolution, to large volume with low resolution, allow to get a robust estimate covering large ranges in terms of mass, volume and redshift. It also allows to test for variations with cosmological parameters, and check for agreements when size and resolution overlap.

The most massive haloes are the hosts of the galaxy clusters. These haloes are sensitive to the initial matter density field and to the growth of structures, and depends on the underlying cosmological model. Galaxy clusters are thus a powerful cosmological probe. The difficulty here is that one needs to link an observable of the galaxy cluster (such as the number of galaxies, the Sunyaev-Zeldovich or X-Ray emission of the gas) to the total mass of the halo in order to fit the HMF and get the cosmology (e.g [Allen et al. 2011](#), for a review).

The whole [Chapter 2](#) is dedicated to the calibration of the relation between dark matter haloes and their host galaxies. We will use halo mass functions to calibrate this relation and track its evolution with time.

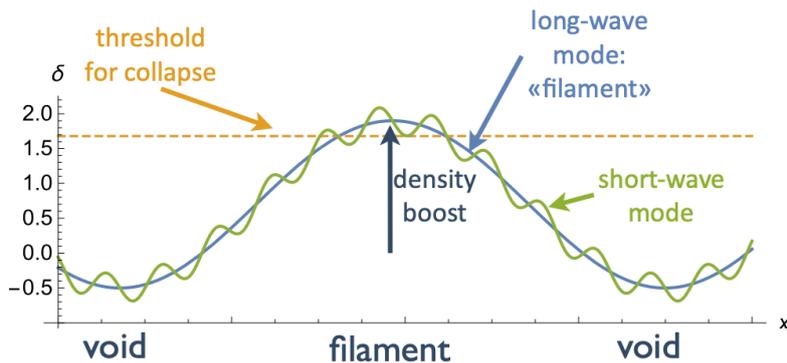


Figure 1.10: Figure from [Codis \(2016\)](#). The combination of large density modes with smaller ones allows density peaks to reach the  $\delta_c$  threshold more easily in dense regions. It is then more probable to form haloes in dense regions. Haloes are thus a biased tracer of the density field.

## 1.7 Astrophysical effects on galaxy power spectrum

We introduce here two main astrophysical effects that impact the power spectrum of galaxies: the galaxy bias and the redshift space distortions (RSD). We need to take these effects into account to accurately link the observed power spectrum of galaxies to the matter power spectrum we derived in the Section 1.5.1.

### 1.7.1 Bias of haloes and galaxies

Both dark matter haloes and galaxies are sampling the matter density field, but they do not sample it uniformly, we observe numerous galaxies in dense environment, and almost no galaxies in less dense environments. We say that haloes, and the galaxies they host, are biased tracers of the underlying matter density field. This idea was first described in [Kaiser \(1984\)](#) and refined in [Bardeen et al. \(1986\)](#).

The threshold of linear density  $\delta_c = 1.686$  seen in Section 1.6.2 explains why haloes are biased tracers of the density field. Due to the presence of both large and short wavelength modes in the density field, a large overdensity inside a void will not collapse as it cannot reach the  $\delta_c$  threshold, while a small overdensity in a very dense region will collapse. As such haloes can collapse more easily in dense regions. The Figure 1.10 illustrate this fact. Haloes are then populated with galaxies, making galaxies another biased tracer of the underlying matter density field.

The galaxy bias is often assumed to be scale independent, which is true at large scales ([McDonald and Roy 2009](#); [Desjacques et al. 2018](#)) and modelled as function of redshift. We can then express the galaxy 3D density field as

$$\delta_g(z, \hat{\mathbf{n}}) = b(z) \delta(z, \hat{\mathbf{n}}), \quad (1.45)$$

with  $\delta_g$  the galaxy density field and  $\delta$  the matter density field. We then get this expression for the power spectrum of galaxies

$$P_g(z, k) = b^2(z) P(z, k), \quad (1.46)$$

with  $P$  the matter power spectrum.

We see here that the bias behaves like a scaling factor which will affect the amplitude of the power spectrum. As such it is highly degenerate with the cosmological parameter which governs the amplitude of the initial power spectrum  $A_s$  (see Equation 1.26).

Equation 1.46 shows a simplified version of the impact of bias on the observed galaxy power spectrum. Cosmological effects such as primordial non-gaussianities, massive neutrinos, dark energy and modified gravity can complicate this picture, and make the bias vary with the scale. Redshift space distortions (RSD), as introduced below, will also impact this relation (see Desjacques et al. 2018, for a review).

### 1.7.2 Redshift space distortions

The observed redshift of galaxies is the sum of the redshift due to the Hubble flow and the peculiar redshift of the galaxy

$$z_{\text{obs}}(z, \hat{\mathbf{n}}) = z + (1 + z) \frac{\mathbf{v}(z, \hat{\mathbf{n}}) \cdot \hat{\mathbf{n}}}{c}, \quad (1.47)$$

where  $\mathbf{v}$  is the peculiar velocity field of galaxies, and  $z$  is the redshift due to the Hubble expansion.

The redshift is our main quantity to get information about the position of the galaxy along the line of sight axis<sup>6</sup>. The difference between the observed redshift and the redshift due to the Hubble expansion will create distortions on the estimated position of galaxies: galaxies moving in our direction will appear closer to us, and galaxies moving away from us will appear further from us. We make the distinction between the *redshift space*, where the position of galaxies is the one given by their observed redshift, and the *real space* which corresponds to the true position of galaxies.

RSD will induce two main effects on the apparent distribution of galaxies. In the linear regime, at large scales, spherical overdensities will appear flattened on the direction perpendicular to the line of sight, due to the fact that galaxies are on average moving towards the center of mass of the overdensity. Inversely, underdense regions will appear elongated along the line of sight. This is known as the Kaiser effect. In the non-linear regime, we speak about the finger of God effect. Galaxies close to the center of mass of a cluster have a high peculiar velocity in random directions. As a result, in the redshift space the cluster will appear elongated along the line of sight.

As shown by Kaiser (1987) 3D power spectrum of galaxies will be anisotropic in the redshift space, due to RSD, while it is isotropic in the real space.

The peculiar velocity field of galaxies  $\mathbf{v}$  is related to the local matter density field, through the linearised continuity equation

$$\frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla \cdot \mathbf{v} = 0, \quad (1.48)$$

<sup>6</sup>One other example is the Tully and Fisher (1977) empirical relation linking the luminosity and the rotation velocity of the galaxy to get an estimate of its distance, but it is less reliable and difficult to measure at high redshift

where we assumed that the velocity of galaxies is an unbiased tracer of the underlying density field of matter (Strauss and Willick 1995).

We introduce the growth rate, which is the derivative of the logarithm of the growth factor

$$f = \frac{d \ln D}{d \ln a} = -(1+z) \frac{1}{D(z)} \frac{dD}{dz}. \quad (1.49)$$

The growth rate is well approximated by  $f(z) = \Omega_m(z)^\gamma$ , with  $\Omega_m(z) = \Omega_{m,0} (1+z)^3 / (H(z)/H_0)^2$ . The  $\Lambda$ CDM model imposes  $\gamma = 0.55$  (Lahav et al. 1991; Linder 2005). Observing any departure from this value will hint towards a deviation from  $\Lambda$ CDM, such as modified gravity models.

We can then get the following expression of the velocity field, in Fourier space (Padmanabhan et al. 2007)

$$\mathbf{v}(\mathbf{k}) = -i \frac{\mathbf{k}}{k^2} f(z) \delta(\mathbf{k}) = -i \frac{\mathbf{k}}{k^2} \beta(z) \delta_g(\mathbf{k}), \quad (1.50)$$

which introduces the redshift distortion parameter  $\beta(z) = f(z)/b_g(z)$ .

At linear order in terms of density and velocity, we can link the density of matter in redshift space  $\delta^s$  to the density of matter in real space to obtain the Kaiser (1987) formula

$$\delta^s(\mathbf{k}) = (1 + f(z) \mu^2) \delta(\mathbf{k}), \quad (1.51)$$

where  $\mu = \cos(\mathbf{k} \cdot \hat{\mathbf{n}})$ .

We see here that we have created an anisotropy in the density field. As such the 3D power spectrum becomes anisotropic, and must then be expressed as a function of the wave number  $k$  but also in terms of  $\mu$ . We have

$$P^s(k, \mu) = (1 + f(z) \mu^2)^2 P(k) \quad (1.52)$$

where  $P^s$  is the matter power spectrum in the redshift space, and  $P$  is the matter power spectrum in real space. Introducing the galaxy bias, we have the galaxy power spectrum in redshift space which is given by

$$P_g^s(k, \mu) = (b(z) + f(z) \mu^2)^2 P(k) = (1 + \beta(z) \mu^2)^2 P_g(k) \quad (1.53)$$

We see that the bias parameter does not affect the term that appears due to the RSD. This is due to the fact that RSD are sensitive to the matter density field, and are as such not sensitive to the galaxy bias.

## Chapter 2

# Galaxy evolution within dark matter haloes

The previous chapter presented the growth of matter density fluctuations from large scales to the beginning of the non-linear regime, which corresponds to  $k \sim 0.2 h \text{ Mpc}^{-1}$  i.e. a scale of  $\sim 5 h^{-1} \text{ Mpc}$ . We will now zoom inside the dark matter haloes and see how baryons evolve and form the galaxies we observe.

Before going more into details, we can make several preliminary comments. The first one is that the galaxy formation process is highly inefficient. Indeed, only about 7% of the baryon content of the universe lays inside galaxies (Shull et al. 2012). The second one is that when comparing the halo mass function obtained from dark matter simulations to the stellar mass function obtained from observations, one can see that they do not have the same shape (see Figure 2.2 or Silk and Mamon 2012). This means that the galaxy formation process depends at least on the mass of the host halo. In this chapter, we will try to address the following question: How do galaxies form and evolve inside the collapsed virialised haloes of dark matter ? To answer this we will link properties of galaxies, such as their mass or star formation rate, to properties of their host dark matter halo.

In a first part, I will present the analysis and the results obtained in Legrand et al. (2019), where we relate the galaxies observed in the COSMOS field with dark matter haloes from simulations. This work is presented in Sections 2.2 to 2.5

I will then present unpublished and ongoing work. In Section 2.6, I will first use a hydrodynamical simulations to connect star formation inside dark matter haloes to its stellar and dark matter mass. In Section 2.7, I will use observations of the gas content in galaxies at high redshift to understand the cycle of baryonic material, from gas to stars, in dark matter haloes.

All these results will allow us to draw a broad picture of the evolution of baryons inside dark matter haloes. In Section 2.8, I will summarize our findings on the evolution of the stellar-to-halo mass relation and of the cold gas. Finally, I will propose cold gas inflows and AGN activity as an answer to the problem of the very efficient galaxy formation in massive haloes at high redshift.

## 2.1 Galaxies and haloes evolution

### 2.1.1 Gas cooling

In the current picture of galaxy formation, galaxies form by accretion of baryons into dark matter virialised haloes (for recent reviews see [Silk et al. 2014](#); [Naab and Ostriker 2017](#)). Pioneering studies found that the gas has to cool efficiently in order to collapse, to fragment and finally to form stars ([Binney 1977](#); [Silk 1977](#); [Rees and Ostriker 1977](#)). In the range of temperatures  $10^4 - 10^7$  K and low-density environments (less than one particle per  $\text{cm}^3$ ) corresponding to the inter-galactic medium, cooling is achieved via radiative cooling (mainly through bremsstrahlung radiation and recombination of ionised Hydrogen and Helium).

The cooling efficiency of a clump of gas is defined by the function  $\Lambda(T, n, Z)$  which depends on the temperature  $T$  of the cloud, its density  $n$  and its metallicity  $Z$ <sup>1</sup> ([Tucker and Gould 1966](#)). One particular feature of gas clouds with such low densities is that the cooling efficiency drops drastically when the cloud temperature reaches  $10^4$  K as all the Hydrogen has become neutral and there are no other efficient cooling processes.

The cooling time is the time it would take for a gas clump inside the halo to radiate all its energy. It is the ratio between the kinetic energy of the gas and the cooling efficiency ([Peacock 1999](#))

$$t_{\text{cool}} \equiv \frac{3 k_{\text{B}} T}{2 \Lambda(T) n}, \quad (2.1)$$

where  $k_{\text{B}}$  is the Boltzmann constant.

This cooling time must be compared with the time for the halo to contract under the gravity, which is given by the free-fall time

$$t_{\text{ff}} = \sqrt{\frac{3\pi}{32 G \rho}}. \quad (2.2)$$

If the cooling time is longer than the free fall time, the cloud is in quasi-static equilibrium: when the temperature decreases due to radiative cooling, the cloud contracts, which increases its density and temperature. In this condition, the cloud cannot collapse as pressure and gravity are always balanced.

If the cooling time is shorter than the free-fall time, then the temperature will drop drastically to  $10^4$  K and the pressure will not be able to compete gravity: the cloud will collapse. The frontier between quasi-static equilibrium and collapse is where  $t_{\text{cool}} \simeq t_{\text{ff}}$ , which is displayed in [Figure 2.1](#). In this Figure we see that galaxies are in the area where the collapse is allowed, while galaxy clusters are still non-collapsed. Indeed observations show us that clusters are filled with hot gas (e.g. [Forman and Jones 1990](#)) and are the hosts of several galaxies, instead of being only one gigantic collapsed cloud of gas forming only one gigantic galaxy.

The metallicity plays an important role here, as cooling efficiency increases when the metallicity is high. The metallicity of the gas increases with time, as massive

---

<sup>1</sup>The metallicity is the mass fraction of all elements more massive than Hydrogen and Helium.

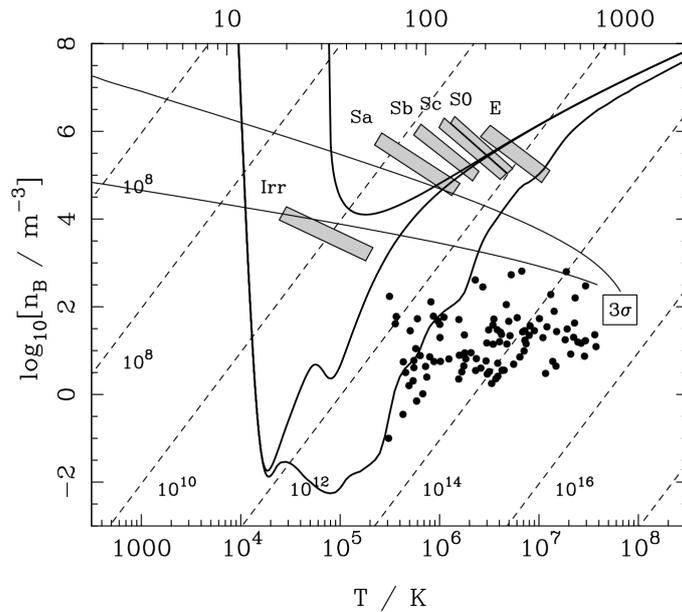


Figure 2.1: Figure from [Peacock \(1999\)](#). Diagram of the baryonic density versus temperature of a gas clump. The diagonal broken lines are lines of constant total mass (in Solar units), assuming 10% of the total mass to be baryonic. The solid curves with minima show the locus along which the cooling time equals the free-fall time; objects above these curves can cool efficiently and collapse. The bottom curve is for a solar metallicity, the middle one is for a zero metallicity and the top one includes the impact of photo-ionization, which makes cooling less efficient. The location of galaxies is shown by the grey boxes, and galaxy clusters are represented by the solid dots. This diagram draws the line between haloes hosting galaxies, where the gas was able to collapse, and clusters filled with hot non-collapsed gas. This line corresponds roughly to haloes of  $M_h = 10^{12} M_\odot$ .

stars create heavy elements and disperse them in the gas with supernovae explosions. This means that late time haloes should be more efficient at cooling than early time haloes. The process of star formation is then more efficient at low redshift than at high redshift. Assuming that the “reservoir” (or supply) of matter is steady, we should see a higher star formation at low redshift than at high redshift.

However, we observe the opposite: star formation increases with redshift, with a peak at  $z \sim 2$  ([Madau and Dickinson 2014](#)). Going from  $z \sim 2$  to today ( $z = 0$ ), the average star formation at the scale of the Universe decreases. We will see in [Section 2.8](#) that this evolution of star formation could be associated to the starvation of cold gas inflows that used to fuel massive galaxies at high redshift.

### 2.1.2 Galaxy formation efficiency

Galaxies form inside dark matter haloes. In the hierarchical clustering scenario, small haloes are accreted onto more massive ones and become “sub-haloes”. Galaxies are classified as either “satellite” if they are hosted by a sub-halo, or “central”. Central galaxies correspond to the most massive galaxy in the halo, located near its center

of mass. For the sake of simplicity, we often refer to any halo or sub-halo hosting a galaxy as a “halo”.

If we make the simple hypothesis that the stellar content of a halo is directly proportional to its mass, the ratio of stellar mass to halo mass would be a constant. However, we do not observe a straightforward linear relation between the dark matter halo mass ( $M_h$ ) and the stellar mass of a galaxy ( $M_*$ ).

This can be seen quantitatively if one compares the dark matter halo mass function and the galaxy stellar mass function, see Figure 2.2. The stellar mass function (SMF) gives the number of galaxies per unit volume and per unit of mass. It has traditionally been modelled by a [Schechter \(1976\)](#) function, although for certain galaxy populations a combination of more than one such functions may provide a better fit to observations ([Binggeli et al. 1988](#); [Kelvin et al. 2014](#)).

In Figure 2.2 where we compare the stellar mass function to the halo mass function, we see that there is a deficit of galaxies with a low stellar mass ( $M_* < 10^{11}M_\odot$ ) and with a high stellar mass ( $M_* > 10^{11}M_\odot$ ), compared to the number density of dark matter haloes (see also e.g. [Cole et al. 2001](#); [Yang et al. 2003](#); [Eke et al. 2006](#); [Behroozi et al. 2010](#); [Moster et al. 2010](#)).

Although such comparisons between mass functions are phenomenological in nature ([Mutch et al. 2013](#)), they provide useful indications to models of galaxy formation in particular when the comparison spans a large redshift range. Two main physical processes have been proposed to explain the discrepancy at low and high mass. Massive galaxies often host Active Galactic Nuclei (AGN) which will heat the gas and expel the gas that would form the stars outside of the galaxy (for a review see [Heckman and Best 2014](#)). Small galaxies instead are not massive enough to retain the gas expelled by supernovae explosions (see e.g. [Larson 1974](#); [Powell et al. 2011](#)) which decreases the available material to form stars. In both cases, these processes are lowering the efficiency of star formation, see [Silk and Mamon 2012](#) for a detailed review on the feedback mechanisms.

### 2.1.3 Stellar to halo mass relation

Understanding how the stellar mass content of a galaxy relates to the mass of its dark matter halo is an alternative way of considering the process of galaxy formation.

Highly complete mass-selected galaxy surveys at high redshift (see, e.g., [Ilbert et al. 2013](#)) and accurate predictions for the halo mass function ([Tinker et al. 2008](#); [Watson et al. 2013](#); [Despali et al. 2016](#)) allows us to measure the stellar-to-halo mass relation (SHMR) of galaxies at different epochs.

There are many techniques to accomplish this. In the following, we will use the “sub-halo abundance matching” technique, where the number density of galaxies (from observations) and dark matter sub-haloes (from simulations) are matched to derive the SHMR at a given redshift (see, e.g., [Marinoni and Hudson 2002](#); [Behroozi et al. 2010, 2013, 2018](#); [Moster et al. 2010, 2013, 2018](#); [Reddick et al. 2013](#)). This technique can also be implemented by assuming a non-parametric monotonic relation between the luminosity or stellar mass of the observed galaxies and sub-halo masses at the time of their infall onto central haloes ([Conroy et al. 2006](#)).

Other studies use a “halo occupation distribution” modeling (HOD, see e.g. [Vale](#)

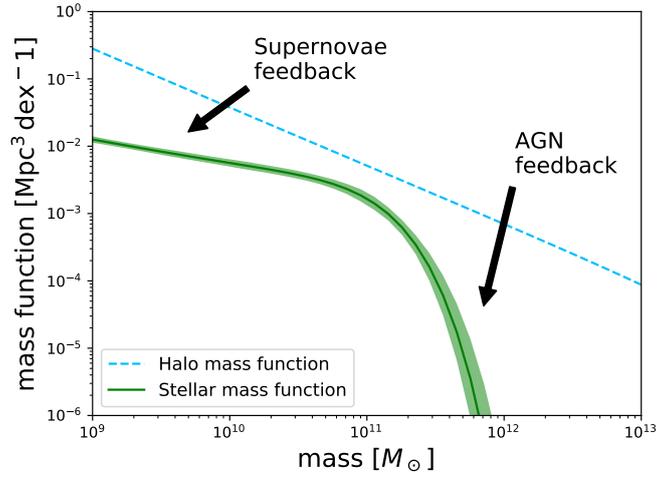


Figure 2.2: Comparison of the stellar mass function from [Davidzon et al. \(2017\)](#) (green line) with the halo mass function from [Despali et al. \(2016\)](#) at a redshift  $z = 0.2$ . The dashed blue line show the halo mass function rescaled by the universal baryon fraction  $\Omega_b/\Omega_m$  to illustrate the maximum available stellar content (assuming all baryons are converted into stars). The shaded green area show the  $1\sigma$  confidence limit of the SMF of [Davidzon et al. \(2017\)](#). It is clear from this Figure that galaxies at low mass ( $M_* < 10^{11}M_\odot$ ) and galaxies at high mass ( $M_* > 10^{11}M_\odot$ ) have a deficit of stellar mass compared to the ones around  $10^{11}M_\odot$ . This can be explained for low mass galaxies by the Supernovae explosions that expels the stellar content out of the galaxy, and for massive galaxies by the AGN feedback which heat the gas and prevent star formation.

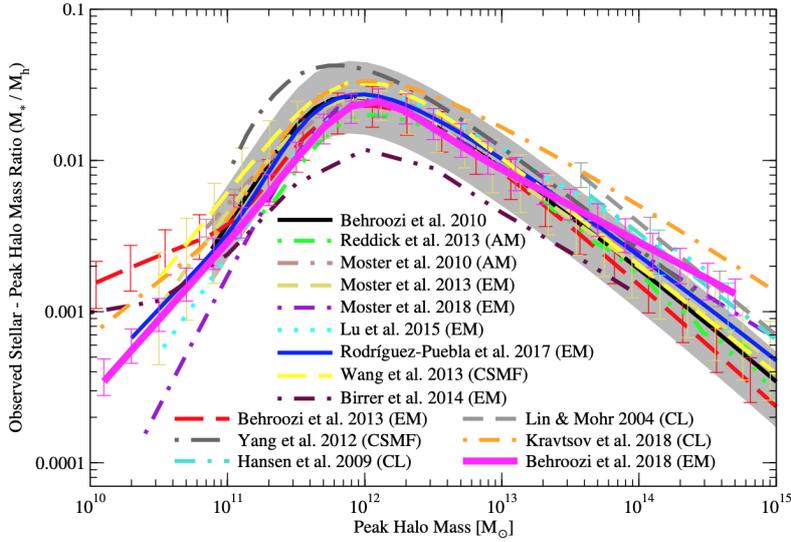


Figure 2.3: Figure from (Behroozi et al. 2018). Compilation of estimations of the stellar to halo mass relation at  $z = 0$  from the literature. This Figure shows the maximal mass in the history of the halo (see Section 2.3.2) versus the stellar to halo mass ratio.

and Ostriker 2004; Zheng et al. 2007; Leauthaud et al. 2011; Coupon et al. 2015) where a prescription for how galaxies populate dark matter haloes can be used to simultaneously predict the number density of galaxies and their spatial distribution. In this case, lensing combined with clustering measurements can provide additional constraints on the SHMR.

Figure 2.3 from Behroozi et al. (2018) shows the ratio of the stellar mass of a galaxy over the host dark matter halo mass for different results from the literature. We see here that there is a “characteristic halo mass”  $M_h^{\text{peak}} \simeq 10^{12} M_\odot$  at which the  $M_*/M_h$  ratio is maximised. A natural interpretation is that  $M_h^{\text{peak}}$  corresponds to the halo mass at which star formation, integrated over the entire assembly history of the galaxy, has been the most efficient (Silk et al. 2014). We consider “galaxy formation efficiency” as the global process of forming stars in dark matter haloes, from the accretion of gas to the actual transformation of baryons into stars. At lower and higher halo masses, the  $M_*/M_h$  ratio decreases rapidly, presumably as a consequence of the physical processes that suppress star formation in these haloes, such as AGN feedback or supernovae explosions (see Section above.)

## 2.2 Galaxies

### 2.2.1 The COSMOS field

Our work is based on the stellar mass function of Davidzon et al. (2017) estimated from observations of the COSMOS field (Scoville et al. 2007b). The COSMOS field displayed in Figure 2.4 is a  $2 \text{ deg}^2$  field with deep UV-to-IR coverage. It is a gathering of several observations of the same patch of the sky, observed with ground based telescopes and satellites such as the *Hubble* Space Telescope or the *Spitzer* Space

Telescope (see [Laigle et al. 2016](#), and references therein).

The large area of COSMOS makes it ideal to collect robust statistics of distant, massive galaxies. Moreover, exquisite IR photometry means precise stellar mass estimates can be made over a large redshift range (see e.g. [Steinhardt et al. 2014](#); [Davidzon et al. 2017](#)). Extensive tests have been made to validate the mass completeness and the photometric redshift accuracy in COSMOS ([Laigle et al. 2016](#); [Davidzon et al. 2017](#)). Far-IR, radio, and X-ray observations are also available to assess the crucial role of AGN ([Delvecchio et al. 2017](#)), and the quenching of distant and massive galaxies ([Gozaliasl et al. 2018](#)).

The COSMOS catalog of [Laigle et al. \(2016\)](#) is a very deep and complete census of more than half a million galaxies up to  $z \sim 6$ . It represents a wealth of information that can be used to understand the formation and evolution of galaxies.

In [Laigle et al. \(2016\)](#), redshifts of more than half a million galaxies out to  $z \sim 6$  were estimated with *Spectral Energy Distribution* (SED) fitting. This method consists in fitting templates of galaxy spectra on observed photometric fluxes in given photometric bands. This is a low resolution spectroscopic redshift estimate. The uncertainty on the photo- $z$  is estimated by comparing to spectroscopic measurements which has been obtained for almost 100 000 galaxies of the catalog ([Lilly et al. 2007](#); [Le Fèvre et al. 2005, 2015](#); [Hasinger et al. 2018](#)).

In [Davidzon et al. \(2017\)](#), new estimates of the photometric redshift and of the stellar mass have been made for galaxies at  $z > 2.5$ , based on an optimized method for  $3 < z < 6$ . This increased the accuracy of the stellar mass function at high redshift. The  $1\sigma$  error on the photometric redshift is estimated to be  $0.03(1+z)$  and the outliers fraction (catastrophic photo- $z$  estimation) is of 12%.

The unique combination of deep optical (Subaru), near-infrared (VISTA) and mid-infrared (*Spitzer*/IRAC) observations results in a galaxy sample that is  $>90\%$  complete for  $M_* > 10^{10} M_\odot$  up to  $z = 4$ ; and it is  $>70\%$  complete for  $M_* > 10^{10} M_\odot$  at  $4 < z < 6$ .

## 2.2.2 Stellar mass function

We use stellar mass functions derived by [Davidzon et al. \(2017, hereafter D17\)](#) for galaxies in the UltraVISTA-Ultra deep region of the COSMOS field (see [McCracken et al. 2012](#)).

D17 estimated the SMF in ten redshift bins from  $z = 0.2$  to  $z = 5.5$  (see [Figure 2.5](#)) using three independent methods: the  $1/V_{\max}$  technique ([Schmidt 1968](#)), the step-wise maximum likelihood ([Efstathiou et al. 1988](#)) and the maximum likelihood method of [Sandage et al. \(1979\)](#). These three estimators provide consistent SMF estimates. However, they are all affected by observational uncertainties ( $M_*$  and  $z_{\text{phot}}$  errors) that scatter galaxies from their original mass bin. This systematic effect, known as [Eddington \(1913\)](#) bias, dominates at high masses ( $M_* \gtrsim 10^{11} M_\odot$ ) because here galaxy number density declines exponentially; this produces an asymmetric scatter and consequently modifies the SMF profile. Depending on the “skewness” and the magnitude dependence of observational errors the Eddington bias may have a strong impact also at lower masses ([Grazian et al. 2015](#)).

When fitting a [Schechter \(1976\)](#) function to their  $1/V_{\max}$  SMF determinations,

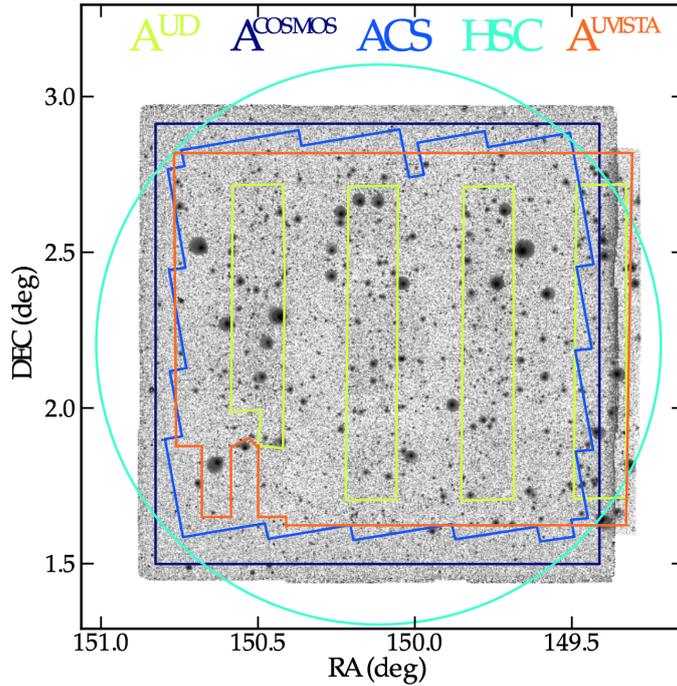


Figure 2.4: Figure from [Laigle et al. \(2016\)](#). The COSMOS field. Colour lines show the different observations footprints.

D17 account for the Eddington bias using the method introduced in [Ilbert et al. \(2013\)](#). Therefore in our work we use these Schechter fits of D17 which should be closer to the intrinsic SMF compared to the other estimators. For consistency, we rescale these estimates to [Planck Collaboration et al. \(2016a, P16\)](#) cosmology. The fitting function assumed by D17 is a double Schechter (see Eq. 4 in D17) at  $z < 3$  and a single Schechter function (their Eq. 3) above that redshift. At low redshifts two SMF components are clearly visible (e.g. [Ilbert et al. 2010](#)), above  $z > 3$  there is no evidence of this double Schechter profile ([Wright et al. 2018](#)). The SMF error bars include both systematic and random errors including Poisson noise, cosmic variance and the scatter due to uncertainties in the SED fitting.

## 2.3 Dark matter haloes

### 2.3.1 Halo mass function

Our main reference for the dark matter halo mass function<sup>2</sup> (HMF) is the work of [Despali et al. \(2016\)](#). They measure the HMF using six  $N$ -body cosmological simulations with different volumes and resolutions: all of them have  $1024^3$  dark matter particles with masses ranging from  $1.94 \times 10^7$  to  $6.35 \times 10^{11} h^{-1} M_{\odot}$  and a corresponding box size from  $62.5$  to  $2000 h^{-1} \text{Mpc}$ . Haloes are identified through the “spherical overdensity” algorithm ([Press and Schechter 1974](#)), i.e. each halo is a sphere with a matter density equal to the virial overdensity (see [Eke et al. 1996](#)) at the given redshift (which is equal to the median  $z$  of the observed SMF, see Table

<sup>2</sup>HMFs were computed using the COLOSSUS python module ([Diemer 2018](#)).

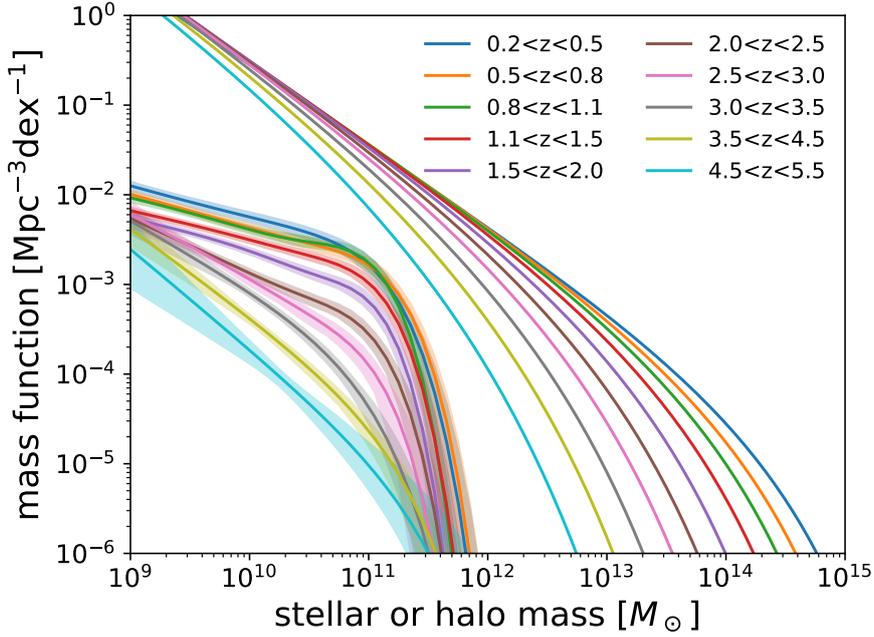


Figure 2.5: Figure from [Legrand et al. \(2019\)](#). Our adopted stellar and halo mass functions. For a given redshift (see legend) the top solid line shows our HMF ([Despali et al. 2016](#), fitted on the Bolshoi-Planck simulation) whilst the solid line and shaded area is the SMF with the associated  $1\sigma$  uncertainty (corresponding to the best fit to  $1/V_{\max}$  points corrected for Eddington bias, [Davidzon et al. 2017](#)).

2.1). The halo mass is defined as the sum of dark matter particles included in such a sphere.

### 2.3.2 Fit on dark matter simulation

It has been shown (see e.g. [Reddick et al. 2013](#)) that for abundance matching applications the stellar mass of galaxies is better correlated to the maximal mass the dark matter haloes have over their history ( $M_{h,\max}$ ) rather than the actual mass at a given redshift. This is particularly true for sub-haloes which can lose mass due to gravitational stripping by the neighbouring main halo whilst the galaxy inside will keep the same stellar mass. Figure 2.6 illustrates the evolution with time of the mass of a halo. [Reddick et al. \(2013\)](#) has demonstrated that using this  $M_{h,\max}$  better fits to observations such as galaxy clustering for abundance matching.

In order to carry out an abundance matching algorithm with the mass  $M_{h,\max}$  instead of the current mass, we fit the [Despali et al. \(2016\)](#) HMF on outputs from the Bolshoi-Planck simulation ([Rodríguez-Puebla et al. 2016](#); [Behroozi et al. 2018](#)).

This dark-matter-only simulation has a comoving volume of  $250h^{-1}\text{Mpc}$  on a side with  $2048^3$  particles with a mass resolution of  $1.6 \times 10^8 h^{-1} M_{\odot}$  and uses [Planck Collaboration et al. \(2016a\)](#) cosmology. Haloes are identified with the ROCKSTAR halo finder and masses are computed using the virial overdensity criterion of [Bryan and Norman \(1998\)](#). [Behroozi et al. \(2018\)](#) provides halo number densities for several halo mass bins and for 178 snapshots from  $z = 16$  to  $z = 0$  for this simulation, with

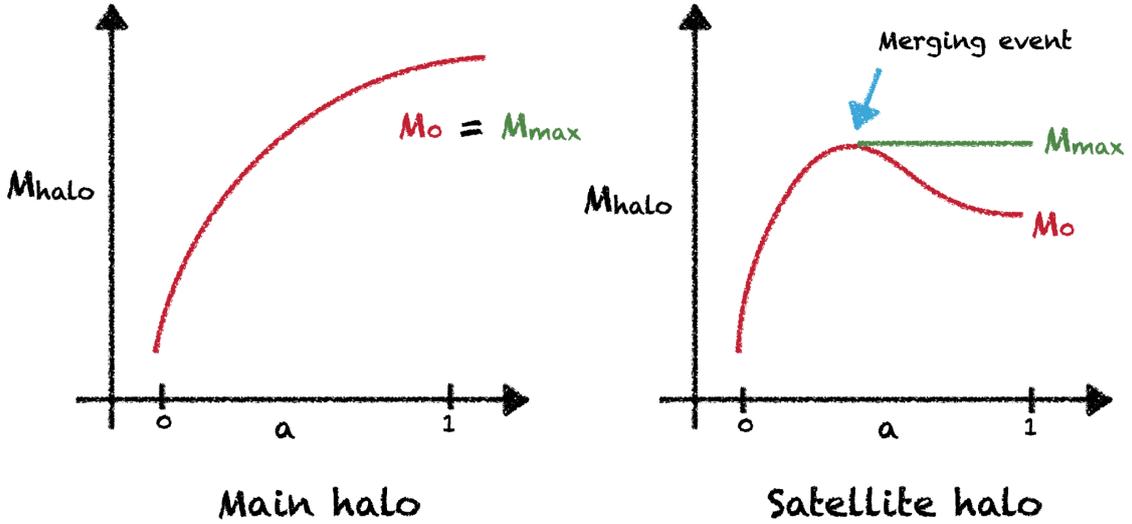


Figure 2.6: Diagram of the evolution of the mass of a given halo with time. Left panel shows the evolution of a main halo which grows by accretion of smaller haloes. Right panel shows the evolution of a halo which became “sub-halo” and lose some mass during the merging event. It has been shown (Reddick et al. 2013) that the maximum mass of the halo  $M_{h,\max}$  (green line) is more correlated to the stellar mass of its host galaxy than the actual mass at a given redshift (red line).

different mass definition, including  $M_{h,\max}$ .

We fit the HMF of Despali et al. (2016) using a modified version of the COLOSSUS code for these data points in the range  $0 < z < 5$  and  $10^{11} h^{-1} M_{\odot} < M_{h,\max} < 10^{15} h^{-1} M_{\odot}$ . The HMF is defined by (Equation 7 of Despali et al. 2016)

$$\nu f(\nu) = A_0 \left( 1 + \frac{1}{\nu'^p} \right) \left( \frac{\nu'}{2\pi} \right)^{1/2} \exp(-\nu'/2), \quad (2.3)$$

$$\nu = \delta_c^2 / \sigma^2(M), \quad (2.4)$$

$$\nu' = a \nu, \quad (2.5)$$

where  $\delta_c(z)$  is the critical linear theory overdensity required for spherical collapse (see Section 1.6.2) divided by the growth factor (Kitayama and Suto 1996), and  $\sigma(M, z)$  is the variance of the fluctuations of density smoothed at a radius  $R = (3M/4\pi/\rho_b)^{1/3}$  with  $\rho_b = \rho_c \Omega_{m,0}$  the mean background matter density.

This equation allows us to fit the HMF at all redshifts with only three free parameters,  $(a, p, A_0)$  which define the high-mass cutoff, the shape at lower masses, and the normalisation of the HMF, respectively. Fitting this equation on the 178 snapshots HMFs using  $M_{h,\max}$  halo masses from the Bolshoi-Planck simulation, we find:  $a = 0.831, p = 0.351, A_0 = 0.331$ . Figure 2.7 shows the resulting HMF for several redshifts from 0 to 5.

## 2.4 Estimating the stellar-to-halo mass relation

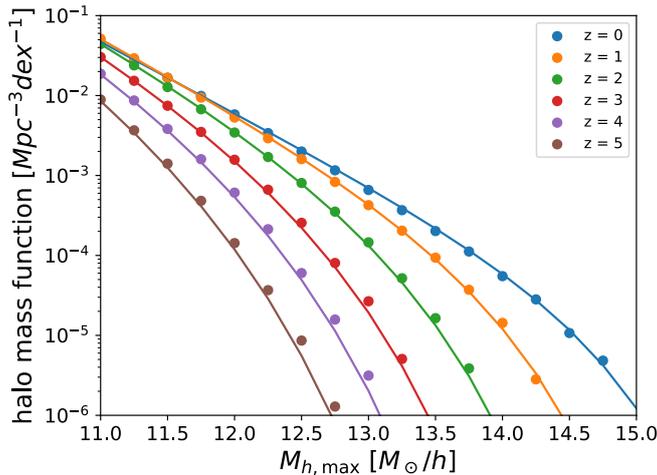


Figure 2.7: Figure from [Legrand et al. \(2019\)](#). Points show halo densities obtained from the Bolshoi-Planck simulation for different redshift snapshots, using the maximal mass in the history of the haloes. The fit of the [Despali et al. \(2016\)](#) HMF on this data points is shown as the plain lines.

### 2.4.1 Implementation

In the sub-halo abundance matching (SHAM) technique, dark matter haloes and galaxies are ranked according to their mass, and then the galaxies are associated to haloes by assuming a monotonic one-to-one relationship ([Vale and Ostriker 2004](#); [Conroy et al. 2006](#); [Behroozi et al. 2010](#); [Moster et al. 2010](#); [Reddick et al. 2013](#)).

Note that in our definition of the halo mass (computed with the spherical overdensity criterion), the mass of the main halo includes the mass of possible sub-haloes, while the mass of the central galaxy does not include the mass of the satellite galaxies. The SHAM method also does not consider either the gas mass or the intracluster medium.

We carry out a “parametric” SHAM, assuming a functional form for the relation between  $M_*$  and  $M_h$ . Following the same formalism as in [Behroozi et al. \(2010\)](#), such a parametric SHMR is described by the following equation:

$$\log(M_h) = \log(M_1) + \beta \log(M_*/M_{*,0}) + \frac{(M_*/M_{*,0})^\delta}{1 + (M_*/M_{*,0})^{-\gamma}} - \frac{1}{2}. \quad (2.6)$$

This model has five free parameters ( $M_1$ ,  $M_{*,0}$ ,  $\beta$ ,  $\delta$ ,  $\gamma$ ) which determine the amplitude, the shape and the knee of the SHMR (see [Behroozi et al. 2010](#), for a more detailed description of the role of each parameter in shaping the SHMR).

The galaxy cumulative number density ( $N_*$ ) and the halo cumulative number

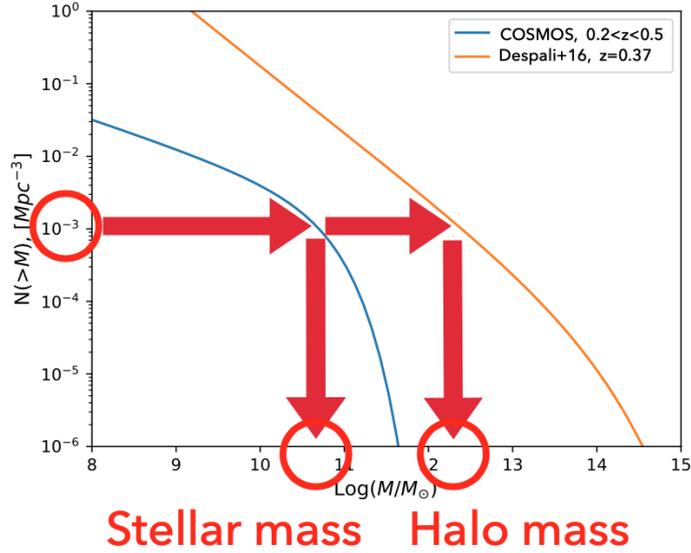


Figure 2.8: Diagram of the abundance matching procedure. The orange line shows the cumulative distribution of the [Despali et al. \(2016\)](#) HMF and the blue line is the cumulative distribution of [Davidzon et al. \(2017\)](#) SMF for the redshift bin  $0.2 < z < 0.5$ . Stellar mass and halo mass are linked by equating  $N_*(M_*) = N_h(M_h)$ .

density ( $N_h$ ) above a certain mass are respectively given by

$$N_*(M_*) = \int_{M_*}^{+\infty} \phi_*(M) dM \quad (2.7)$$

$$N_h(M_h) = \int_{M_h}^{+\infty} \phi_h(M) dM \quad (2.8)$$

with  $\phi_*$  and  $\phi_h$  being the stellar and halo mass functions. The main assumptions of SHAM is that there is only one galaxy per dark matter halo and that the relation between stellar and halo masses is monotonic. As a consequence, the  $M_*$  value associated to a given  $M_h$  is the one for which  $N_*(M_*) = N_h(M_h)$ . This procedure is illustrated in [Figure 2.8](#).

The derivative of this equation gives the relation between SMF, HMF, and SHMR:

$$\phi_{*,\text{conv}}(M_*) = \frac{dM_h}{dM_*} \phi_h(M_h), \quad (2.9)$$

where the differential term on the right-hand side can be derived from [Equation 2.6](#).

We use the notation  $\phi_{*,\text{conv}}$  because we multiply the SMF  $\phi_*$  with a log-normal distribution with a standard deviation  $\xi$  to account for scatter in stellar mass at fixed halo mass.

The standard deviation ( $\xi$ ) of this log-normal distribution is kept as an additional free parameter; we assume that  $\xi$  is independent of the halo mass ([More et al. 2009](#); [Moster et al. 2010](#)) but can vary with redshift. We note here that new hydrodynamical simulations like Eagle ([Schaye et al. 2015](#)) have shown that this scatter decreases from 0.25 dex at  $M_h = 10^{11} M_\odot$  to 0.12 dex at  $M_h = 10^{13} M_\odot$  (see [Matthee et al. 2017](#)). This evolution of the scatter is in agreement with latest abundance

matching models (Coupon et al. 2015; Behroozi et al. 2018; Moster et al. 2018). See also figure 9 of Gozaliasl et al. (2018). However in our analysis, we restrict to a mass-invariant scatter for simplicity.

The model SMF defined in Eq. (2.9) is then fitted to the observed one (i.e.,  $\phi_{*,\text{obs}}$ ) through the procedure described in the next Section.

### 2.4.2 Fitting procedure

To fit the model SMF to real data, we define the negative log-likelihood

$$\chi^2 = \sum_i \left( \frac{\phi_{*,\text{conv}}(M_{*,i}) - \phi_{*,\text{obs}}(M_{*,i})}{\sigma_{\text{obs}}(M_{*,i})} \right)^2, \quad (2.10)$$

where  $\sigma_{\text{obs}}$  is the uncertainty of the observed SMF in a given stellar mass bin  $M_{*,i}$  (with the first bin starting at  $M_{*,\text{min}}$ ).

For each of the ten redshift bins, we minimise Equation (2.10) using a Markov Chain Monte Carlo (MCMC) algorithm<sup>3</sup>. This algorithm allows the sampling of the parameter space in order to derive the posterior distribution for the six free parameters. We use flat conservative priors on the parameters together with 250 walkers each with a different starting point randomly selected in a Gaussian distribution around the original starting point. The convergence criterion is based on the autocorrelation length, which is an estimate of the number of steps between which two positions of the walkers are considered uncorrelated (Goodman and Weare 2010).

The MCMC stops when the autocorrelation length has changed by less than 1% and when the length of the chain is at least 50 times the autocorrelation length. As an example, the chains in the case of the HMF fitted on Bolshoi-Planck have a length between 5000 at low redshift and 25000 in the highest redshift bin. With our 250 walkers this gives between  $1.25 \times 10^6$  and  $6.25 \times 10^6$  samples. The first steps up to two times the autocorrelation length are discarded as a burn-in phase. To speed up the computation of the posteriors, we keep only the iterations separated by a thin length which is half of the autocorrelation length.

We show the best fit and the 68% confidence interval for the six free parameters in each of the ten redshift bins in Table 2.1. We show the  $1\sigma$  confidence interval for the redshift bin [0.8, 1.1] in Figure 2.18 (at the end of the chapter). The  $1\sigma$  confidence intervals of each of our ten fits are shown in the Appendix of (Legrand et al. 2019). These figures show that the parameters  $M_1$  and  $M_{*,0}$  are highly correlated. This is expected because as  $M_1$  increases,  $M_{*,0}$  should also increase.  $M_1$  and  $\beta$  are also highly correlated which may be explained by the fact that  $\log(M_*/M_{*,0})$  is negative for a large range of stellar masses so an increase of  $\beta$  is compensated by an increase of  $M_1$ . The value of  $\delta$  is not well constrained at high redshift, because this parameter controls the high mass slope which is not well measured with our data.

<sup>3</sup>We use the EMCEE python package (Foreman-Mackey et al. 2013).

### 2.4.3 Main sources of SHAM uncertainties

There are several sources of uncertainties involved in the SHAM technique. A subhalo may be stripped after infall, leaving the hosted galaxy embedded in the larger, central halo. This may break the one-to-one correspondence between galaxies and dark matter haloes which is the main assumption of our method. The HOD model is a viable solution to take this into account although it would introduce an additional number of assumptions and free parameters. Moreover, observed galaxy clustering is required to constrain the HOD model parameters (e.g. [Coupon et al. 2015](#)) but such measurements are challenging at  $z > 2$  ([Durkalec et al. 2015](#)). At lower redshift ( $z \lesssim 1$ ) [Leauthaud et al. \(2012, see their Figure 13\)](#) have shown that  $M_h^{\text{peak}}$  measurements are consistent between HOD and SHAM measurements.

Another source of uncertainty comes from random and systematic errors in  $z_{\text{phot}}$  and  $M_*$  estimates, with the former propagating into the  $M_*$  error in a way difficult to model (see discussion in D17). In D17, the logarithmic stellar mass uncertainty is described by a Gaussian with standard deviation  $\sigma_{M_*} = 0.35$  dex multiplied by a Lorentzian function with a parameter  $\tau$  increasing with redshift to enhance the tails of the distribution (see equation 1 of D17). These observational uncertainties which cause the Eddington bias have been corrected for in the SMF estimates we adopt (see Section 2.2.2) but some caveats remain (see D17; [Grazian et al. 2015](#)). Moreover in our fitting procedure, we consider that different stellar mass bins are uncorrelated (Eq. 2.10). This assumption is a consequence of the fact that in D17 (as in the vast majority of the literature) covariance matrices are not provided for the SMF estimates. Once corrected for  $M_*$  *observational* uncertainties, the main source of correlations between mass bins comes from the *intrinsic* covariance between them. To avoid oversampling, we adopt a mass bin size of 0.3 dex which is comparable to the scatter in D17. We verified that this choice does not introduce any significant bias: modifying the bin size and centre (by  $\pm 0.1$  dex) results remain consistent within  $1\sigma$ .

A final source of uncertainty comes from the  $M_*$  scatter that we add to the galaxy-to-halo monotonic relation. This is modelled with a log-normal distribution characterised by the parameter  $\xi$  which is free to vary in the MCMC fit. This parameter is usually fixed between 0.15 and 0.20 dex (see e.g. [More et al. 2009](#); [Moster et al. 2010](#); [Reddick et al. 2013](#)) but in the large redshift range probed here we expect a non-negligible variation due to the evolution of galaxies' physical properties as well as observational effects. We note however that the resulting values (see Table 2.1) are compatible with the fixed ones assumed by the studies mentioned previously.

## 2.5 Results: stellar-to-halo mass relation in COSMOS

### 2.5.1 Stellar-to-halo mass ratio

Figure 2.9 shows our derived stellar-to-halo mass ratios. The SHMR and the corresponding  $1\sigma$  uncertainty are computed respectively as the 50<sup>th</sup>, the 16<sup>th</sup> and the 84<sup>th</sup> percentile of the distribution of  $M_h$  at a given  $M_*$  in the remaining MCMC

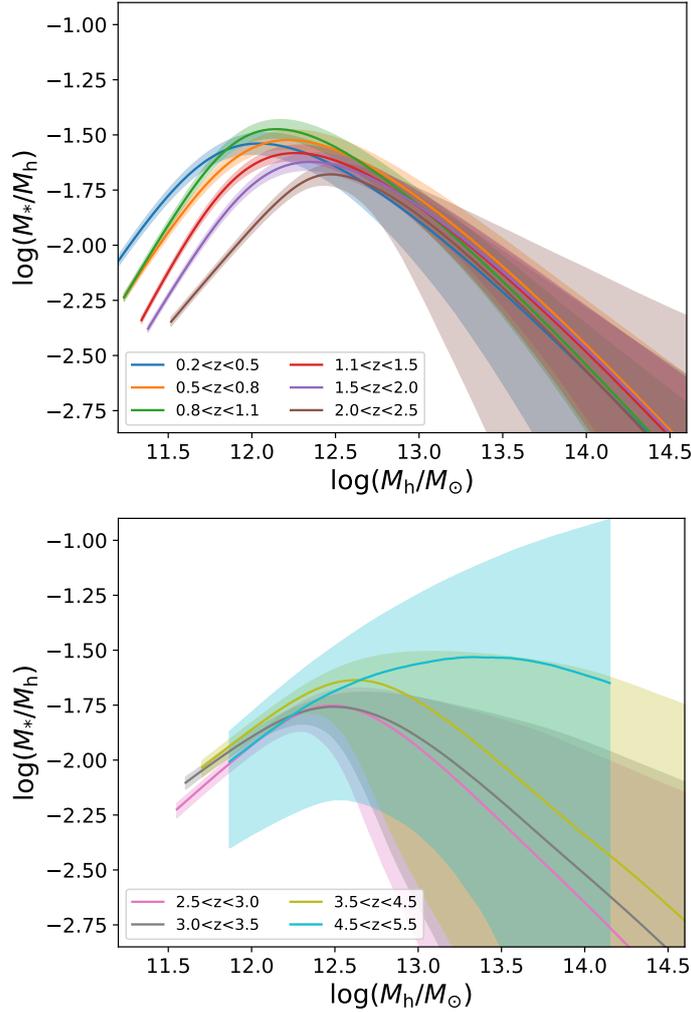


Figure 2.9: Stellar-to-halo mass ratio from  $z = 0.2$  to  $z = 2.5$  (upper panel) and from  $z = 2.5$  to  $z = 5.5$  (lower panel). Thick lines show the 50<sup>th</sup> percentile of the  $M_h$  distribution at fixed  $M_*$  computed from our MCMC runs. The coloured bands show the 16<sup>th</sup> and 84<sup>th</sup> percentile.

chains (Section 2.4.2). These uncertainties are shown as shaded regions.

For each redshift, the  $M_*/M_h$  ratio peaks at  $M_h \simeq M_h^{\text{peak}}$  and drops by one order of magnitude at both the extremes of our halo mass range. At  $z < 0.5$ ,  $M_h^{\text{peak}} = 10^{12} M_\odot$ , with  $\log(M_*/M_h) = -1.55 \pm 0.5$ . At higher redshifts,  $M_h^{\text{peak}}$  increases steadily up to  $10^{12.5} M_\odot$  at  $z = 2$ , i.e. growing by a factor  $\sim 3$ . It then remains flat up to  $z = 4$ . At a fixed halo mass above  $M_h^{\text{peak}}$ ,  $M_*/M_h$  does not evolve, while in haloes below  $M_h^{\text{peak}}$  the ratio decreases from  $z \sim 0$  to  $z \sim 2.5$ .

### 2.5.2 Variation of the peak halo mass with redshift

Figure 2.10 shows the redshift evolution of the peak halo mass between  $z = 0.2$  and 4.5 computed from the median  $M_h^{\text{peak}}$  for all the samples retained in the MCMC (see Section 2.4.2). The results are reported also in Table 2.1. Figure 2.10 also presents a compilation of recent measurements from the literature together with model pre-

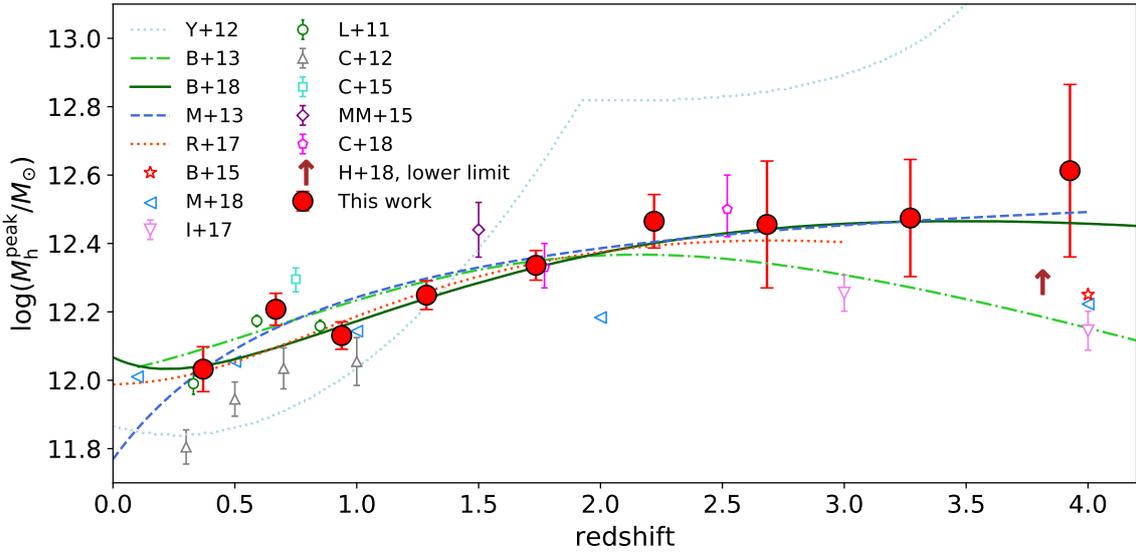


Figure 2.10: Figure from [Legrand et al. \(2019\)](#). Peak halo mass  $M_h^{\text{peak}}$  as a function of redshift (red dots). We plot  $M_h^{\text{peak}}$  at the median redshift of each bin, rescaled to  $H_0 = 70 \text{ km.s}^{-1}\text{Mpc}^{-1}$ . All masses from other studies have been rescaled to match the  $H_0 = 70 \text{ km.s}^{-1}\text{Mpc}^{-1}$  cosmology. Some points from the literature have been slightly shifted along the redshift axis for clarity. We show results from [Leauthaud et al. \(2011, L+11\)](#), [Yang et al. \(2012, Y+12\)](#), [Coupon et al. \(2012, C+12\)](#), [Moster et al. \(2013, M+13\)](#), [Behroozi et al. \(2013, B+13\)](#), [Behroozi and Silk \(2015, B+15\)](#), [Coupon et al. \(2015, C+15\)](#), [Martinez-Manso et al. \(2015, MM+15\)](#), [Rodríguez-Puebla et al. \(2017, R+17\)](#), [Ishikawa et al. \(2017, I+17\)](#), [Cowley et al. \(2018, C+18\)](#), [Harikane et al. \(2018, H+18\)](#), [Moster et al. \(2018, M+18\)](#), and [Behroozi et al. \(2018, B+18\)](#). The brown arrow is the lower limit for  $M_h^{\text{peak}}$  from [Harikane et al. \(2018, H+18\)](#).

dictions (lines). At  $z > 3$  it becomes progressively more difficult to measure the position of the peak as the slopes of halo and stellar mass functions become similar (Figure 2.5). In addition at higher redshifts, there are correspondingly smaller numbers of massive galaxies in the COSMOS volume. Nevertheless, our measurements show clearly that the peak halo mass increases steadily from  $10^{12} M_\odot$  at  $z = 0.3$  to  $10^{12.6} M_\odot$  at  $z = 4$ .

Below  $z \sim 2.5$  there is generally a good agreement in the literature with  $M_h^{\text{peak}}$  steadily increasing as a function of redshift. We confirm this trend despite some fluctuation (e.g., at  $z \sim 0.7$ ) due to the over-abundance of rich structures in COSMOS (see e.g. [McCracken et al. 2015](#)).

Above  $z \gtrsim 3$  the scatter in  $M_h^{\text{peak}}$  increases. [Moster et al. \(2013\)](#) and [Behroozi et al. \(2013\)](#) find different trends, i.e. a  $M_h^{\text{peak}}(z)$  function that declines ([Behroozi et al. 2013](#)) or flattens ([Moster et al. 2013](#)) with increasing redshift. One possible explanation for the discrepancy is that [Moster et al.](#) and [Behroozi et al.](#) models are based on different observational datasets. To address this issue, [Behroozi and Silk \(2015\)](#) repeated [Behroozi et al.](#)'s analysis removing  $z > 5$  constraints (which in their method influence also the fit at lower  $z$ ). However, this test is inconclusive as their  $M_h^{\text{peak}}$  estimate (shown as the star symbol in Figure 2.10)

falls between these curves <sup>4</sup>.

Our higher  $M_h^{\text{peak}}$  values with respect to [Ishikawa et al. \(2017\)](#) and [Harikane et al. \(2018\)](#) may be a consequence of the near-infrared selection (a good proxy for stellar mass, see D17). [Ishikawa et al. \(2017\)](#) and [Harikane et al. \(2018\)](#) samples are selected in rest-frame UV (and a conversion to stellar mass is made through an average  $L_{\text{UV}}-M_*$  relation). Moreover, their redshift estimation is derived (instead of  $z_{\text{phot}}$  estimates) from a Lyman-break colour–colour selection which may result in lower levels of purity and completeness at  $z \sim 3$  ([Duncan et al. 2014](#)).

Recently, revised versions of [Behroozi et al. \(2013\)](#) and of [Moster et al. \(2013\)](#) have been presented in [Behroozi et al. \(2018\)](#) and [Moster et al. \(2018\)](#). This new analysis differs from the former ones by following closely the evolution of individual halo-galaxy pairs through time. This results in a better understanding of the scatter of the SHMR, because this scatter results from the evolution of each halo-galaxy pairs, it is not an arbitrary scatter parameter added to the model. In [Behroozi et al. \(2018\)](#), the feedback model regulating star formation has significantly changed since [Behroozi et al. \(2013\)](#). In the updated model, the  $M_h$  threshold at which 50 per cent of the hosting galaxies are quiescent grows from  $10^{12} M_\odot$  at  $z < 1$  up to  $\sim 10^{13} M_\odot$  at  $z = 3.5$  (see [Figure 2.17](#) and [Section 2.8](#)). As a consequence, the  $M_h^{\text{peak}}$  evolution is now in excellent agreement with both [Moster et al. \(2013\)](#) and our estimates. [Moster et al. \(2018\)](#) peak halo mass shown here corresponds to the peak in the ratio between stellar mass and baryonic mass of galaxies (the  $[M_*/M_b](M_h)$  relation). We assumed here that the ratio between baryonic mass and halo mass is a constant (equal to the universal baryon fraction), giving the same value for the peak halo mass of the  $[M_*/M_b](M_h)$  relation. The difference with our results might be explained by a dependence of the baryon fraction of haloes with mass (see [Kravtsov et al. 2005](#); [Davies et al. 2019](#)).

### 2.5.3 Impact of the halo mass function on our results

#### Halo mass functions from the literature

In order to estimate quantitatively how the choice of HMF impacts the results, we repeat the analysis using different HMFs. Besides the variations due to different cosmological parameters ([Angulo and White 2010](#)), it is difficult to model the uncertainties affecting the HMF. Studies have investigated the impact of different halo finding algorithms which produce changes in the HMF of the order of  $\sim 10\%$  ([Knebe et al. 2011](#)). The HMF estimation will also depend on the simulations used to calibrate it, such as its volume or its resolution. Another potential issue is the impact of baryons (not implemented in [Despali et al.](#)) on the growth of dark matter haloes: [Bocquet et al. \(2016\)](#) show that in hydrodynamical simulations the halo number density decreases by  $\sim 15\%$  at  $z \lesssim 0.5$  with respect to dark matter only, whereas at higher redshift the impact of baryons is negligible.

---

<sup>4</sup> $M_h^{\text{peak}}(z)$  error bars are not explicitly quoted either in [Behroozi et al. \(2013\)](#) or [Moster et al. \(2013\)](#). However, we can quantify them through the uncertainties of their SHMR models. For example in the model of [Moster et al.](#), the  $1\sigma$  confidence level of the  $M_1(z)$  parameter can be used as a proxy, leading to  $M_h^{\text{peak}}$  error bars of the same order of magnitude of ours.

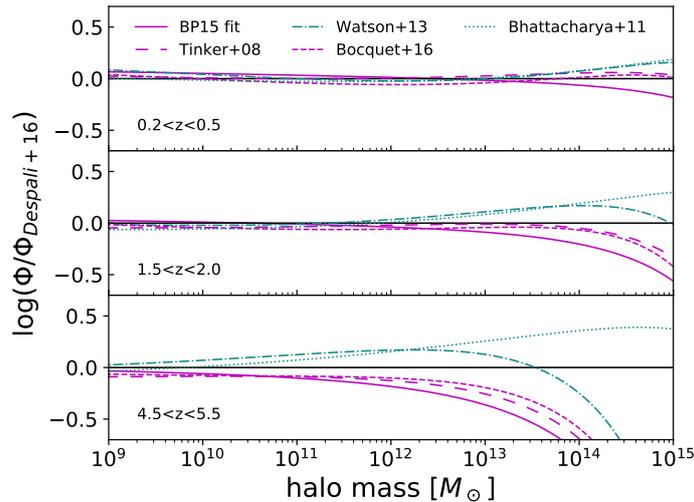


Figure 2.11: Figure from [Legrand et al. \(2019\)](#). Comparison of different halo mass functions from the literature. For three sample redshift bins the relative difference as a function of halo mass between the original [Despali et al. \(2016\)](#) using the virial overdensity criterion, our Bolshoi-Planck fit (solid magenta line), and a selection of HMFs from the literature. Magenta lines show numerical simulations in which haloes are defined according to a spherical over-density threshold (solid line: Bolshoi-Planck, long-dashed line: [Tinker et al. 2008](#), short dashed: [Bocquet et al. 2016](#)). Cyan lines show works that use a friends-of-friends algorithm (dotted line: [Bhattacharya et al. 2011](#), dot-dashed: [Watson et al. 2013](#)).

Our main analysis uses the [Despali et al.](#) mass function fitted on the Bolshoi-Planck simulation as presented in Section 2.3.1. We also compare to the original version of [Despali et al. \(2016\)](#) HMF with halo mass defined using the virial overdensity criterion.

We considered alternate HMF from the literature. These other versions are divided into two categories according to how haloes are identified. HMF estimates in the first category ([Tinker et al. 2008](#); [Bocquet et al. 2016](#)) use the spherical overdensity definition, while the others ([Bhattacharya et al. 2011](#); [Watson et al. 2013](#)) rely on the so-called “friends of friends” (FoF<sup>5</sup>) algorithm ([Davis et al. 1985](#)).

Figure 2.11 shows how the [Despali et al. \(2016\)](#) fit and the other HMF differ from [Despali et al.](#) in three redshift bins. At low redshifts, we find that differences are negligible, in agreement with the literature. However for  $>10^{13} M_{\odot}$  haloes at  $z > 2$ , i.e. in a range barely investigated in previous work, there are 0.2 – 0.5 dex offsets between [Despali et al. \(2016\)](#) and other HMF estimates. Such a difference may be fully explained by Poisson scatter since such massive haloes are rare in the volume of cosmological simulations. We do not attempt to find the physical reasons of the discrepancy and here we simply take the “inter-publication” bias as a measure of HMF uncertainties.

<sup>5</sup>In the FoF algorithm, particles of the simulation are linked by pairs that are closer than a given linking length. Haloes are identified as a combination of particles that are related by a chain of pairs, i.e. they are “friends of friends”.

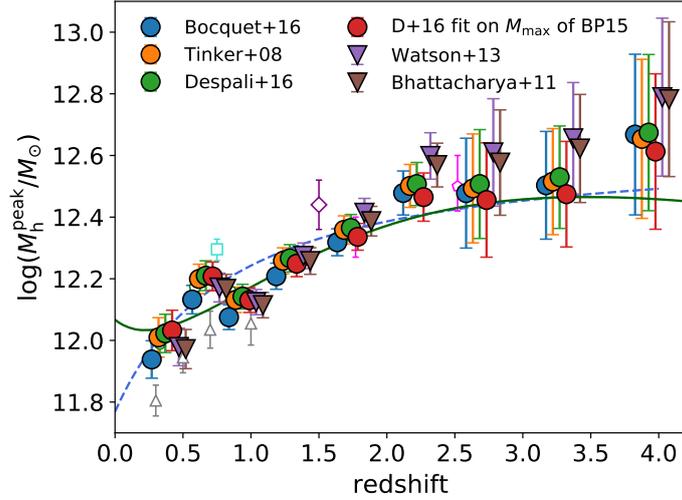


Figure 2.12: Figure from [Legrand et al. \(2019\)](#). Peak halo mass ( $M_h^{\text{peak}}$ ) computed using different HMFs.  $M_h^{\text{peak}}$  redshift evolution is independently measured six times, using different HMF fits: our Bolshoi-Planck fit of [Despali et al. \(2016\)](#), our main reference also shown in [Figure 2.10](#)); the original [Despali et al. \(2016\)](#); [Tinker et al. \(2008\)](#); [Bhattacharya et al. \(2011\)](#); [Watson et al. \(2013\)](#); [Bocquet et al. \(2016\)](#). Filled circles (triangles) indicate that the halo identification has been done with a spherical overdensity (friends-of-friends) algorithm. Each set of  $M_h^{\text{peak}}(z)$  values derived for a given HMF is shifted by 0.05 in redshift for sake of clarity. SHAM method and observed SMF are the same for all estimates. Literature measurements are shown as in [Figure 2.10](#).

### Impact of the halo mass function on $M_h^{\text{peak}}$

We show in [Figure 2.12](#) the impact of changing the HMF on the estimation of the peak halo mass  $M_h^{\text{peak}}$ . Results at  $z \lesssim 2$  are consistent, whilst at higher redshifts we clearly observe the impact of halo identification techniques.  $M_h^{\text{peak}}$  values using the HMF of either [Tinker et al. \(2008\)](#), [Bocquet et al. \(2016\)](#), or [Despali et al. \(2016\)](#) are grouped together, as those studies all applied a spherical overdensity criterion to define haloes. [Bhattacharya et al. \(2011\)](#) and [Watson et al. \(2013\)](#) use a friends-of-friends algorithm, and the resulting  $\log(M_h^{\text{peak}}/M_\odot)$  is systematically higher by  $\sim 0.1$  dex at  $z \geq 2$ . In our study, these differences are smaller than other sources of uncertainties, but it is clear that in future larger surveys these differences may become important.

### 2.5.4 Stellar to halo mass ratio evolution at fixed halo mass

As we can see in [Figure 2.9](#), the value of the stellar to halo mass ratio  $M_*/M_h$  seems to be constant with redshift for massive haloes, while this ratio seems to be increasing with redshift with lower mass haloes.

In order to get a better idea of the galaxy formation efficiency inside dark matter haloes, we show in [Figure 2.13](#) the evolution of the  $M_*/M_h$  ratio with redshift for fixed values of halo mass. We restrict this analysis to  $z < 2.5$  because at high mass

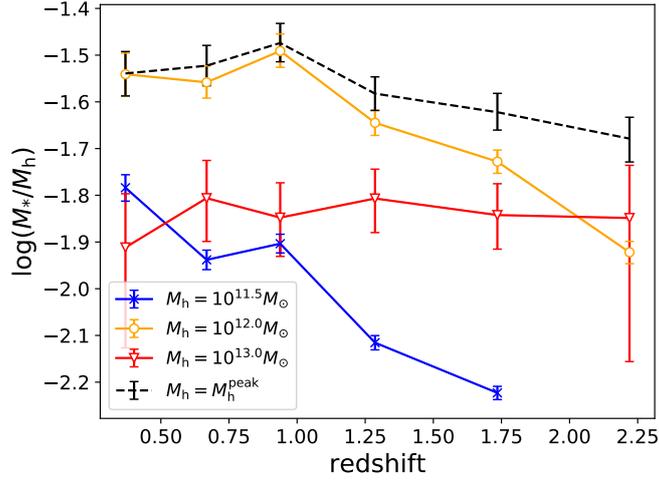


Figure 2.13: Figure from [Legrand et al. \(2019\)](#), showing  $M_*/M_h$  as a function of redshift for fixed halo masses (solid lines). We also show the value of this ratio at the halo mass  $M_h \equiv M_h^{\text{peak}}(z)$  which evolves with redshift (dashed line). Error bars are derived from Figure 2.9.

bins ( $10^{13} M_\odot$ ) uncertainties in the  $M_*/M_h$  ratio prevents a quantitative discussion of its evolution with redshift between  $z = 2.5$  and  $z = 5.5$ .

For massive haloes ( $M_h \simeq 10^{13} M_\odot$ ), the ratio is nearly constant between  $z \sim 0.2$  and 2.5, there is no specific hints of star formation in the halo. For  $M_h \simeq 10^{12} M_\odot$  the ratio increases with cosmic time, and it reaches its maximum value (about 0.03) at  $z \simeq 1$  and then remain constant until  $z \sim 0.2$ .

Lower-mass haloes, which are  $< M_h^{\text{peak}}$  across the whole redshift range, steadily increase their  $M_*/M_h$  without any peak or plateau. For instance, haloes with  $M_h \simeq 10^{11.5} M_\odot$  increase their  $M_*/M_h$  ratio by a factor  $\sim 3.2$  from  $z = 2.5$  to 0.2.

For comparison, Figure 2.13 also shows the increase of the  $M_*/M_h$  ratio, from  $z = 2.5$  to 0.2, for haloes in a mass bin that evolves with redshift, i.e.  $M_h = M_h^{\text{peak}}(z)$ .

### 2.5.5 Interpretation of this evolution

The  $M_*/M_h$  ratio is usually interpreted as the comparison between the amount of star formation and dark matter accretion integrated over a halo’s lifetime. Thus, a high  $M_*/M_h$  ratio in a given  $M_h$  bin implies that those haloes have been (on average) particularly efficient in forming stars. In addition to the *in situ* star formation, further stellar mass can be accreted via galaxy merging. In such a framework, the dependence of the  $M_*/M_h$  ratio on halo mass and redshift can be explained by a combination of physical phenomena (see Section 2.8). Our observational constraint on  $M_h^{\text{peak}}$  can help to understand which mechanisms, amongst those proposed in the literature, are mostly responsible for regulating galaxy stellar mass assembly.

$M_h^{\text{peak}}$  can be considered as the threshold above which haloes maintain a nearly constant  $M_*/M_h$  ratio across time (Figure 2.13). At a fixed halo mass below  $M_h^{\text{peak}}$ , the  $M_*/M_h$  ratio increases with cosmic time, indicating that stellar mass has “kept up” with dark matter accretion. For a fixed halo mass above  $M_h \simeq M_h^{\text{peak}}$ , host

galaxies are more likely to enter in a quiescent phase and thereafter passively evolve. Figure 2.13 clearly shows this evolution with redshift for fixed halo masses. For objects with  $M_h = 10^{12} M_\odot$ , their  $M_*/M_h$  increases until  $z \sim 1$  (i.e., when  $M_h^{\text{peak}} = 1.3 - 1.6 \times 10^{12} M_\odot$ ) after which the ratio remains constant until  $z = 0.2$ . Note that we do not track the evolution of individual haloes but instead the evolution of the  $M_*/M_h$  ratio for a given halo mass. This makes the interpretation of the evolution of individual haloes with time more difficult (haloes at high redshift are not necessarily the same as haloes of the same mass at low redshift).

## 2.6 SHMR from a hydrodynamical simulation

To test the link between  $M_h^{\text{peak}}$  and the efficiency of galaxy formation, we investigate the stellar to halo mass relation in the Horizon-AGN (Dubois et al. 2014) hydrodynamical simulation. We use the lightcone catalog (Laigle et al. 2019; Davidzon et al. 2019) which reproduces the observed properties of galaxies as if they were observed in the same photometric bands and with a similar depth as COSMOS. Dark matter haloes and galaxies are matched in the simulation, allowing us to compare the stellar mass of the central galaxy to the dark matter mass of the halo. The interest of this catalog is that we can get information on the properties of the galaxies that have been measured in the lightcone. In our test case, we are interested in the specific star formation rate (sSFR) of the central galaxy, which is the star formation rate (SFR) (expressed in units of stellar mass per unit of time) divided by the stellar mass of the galaxy.

We fit the stellar-to-halo mass relation of Equation 2.6 on the matched halo-central galaxy catalog. Figure 2.14 shows the average sSFR of galaxies as a function of their stellar mass and halo mass, for different redshift bin of the snapshot. The fitted stellar-to-halo mass relation are the plain black lines and the peak of the  $M_*/M_h$  ratio is shown as the dotted lines. We note that the  $M_h^{\text{peak}}$  values from the simulation are  $\sim 0.5$  dex lower than our measurements in the COSMOS field.

As we can see in Figure 2.14, for a fixed halo mass, there is a scatter of stellar mass around the average value given by the stellar-to-halo mass relation. For a fixed halo mass  $< M_h^{\text{peak}}$ , the sSFR decreases when the stellar mass increase. For a halo mass  $> M_h^{\text{peak}}$ , we see that the sSFR significantly drops compared to lower mass haloes. Alternatively for a fixed stellar mass, the sSFR decreases when the halo mass increases.

This shows that the galaxy formation is more efficient in small haloes hosting a small galaxy. As time goes, halo mass and stellar mass increase and the galaxy formation efficiency decreases. This Figure illustrates the fact that the sSFR is tightly correlated with the position of the galaxy with respect to the stellar-to-halo mass relation, and there is a drop of the star formation of the central galaxy in haloes above the threshold given by  $M_h^{\text{peak}}$ .

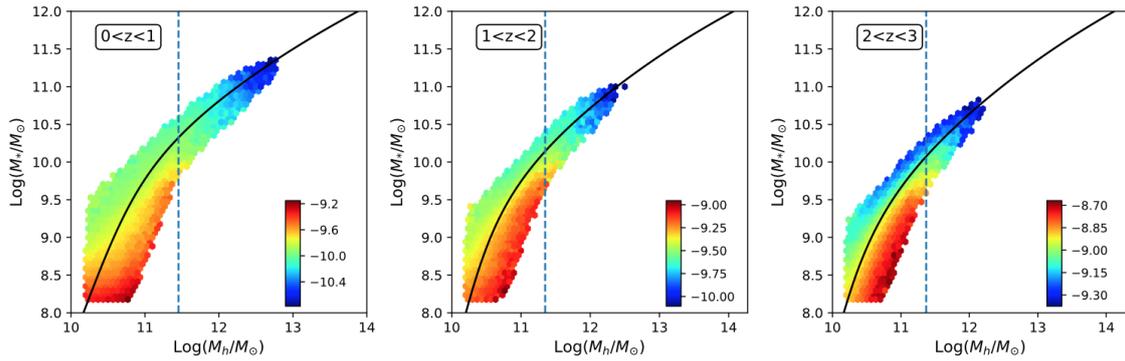


Figure 2.14: Stellar to halo mass relation in Horizon AGN. The hexagonal bins are colour coded as a function of the average specific star formation rate of galaxies that fall in this bin. Panels from left to right are for different redshift bins (shown on the upper left corner of each panel). In each panel, colorbars show the corresponding values of  $\log_{10}(\text{sSFR})$ . The plain line is the fitted SHMR from Equation 2.6, and the vertical dashed line show the  $M_h^{\text{peak}}$  value. We only show bins containing more than 50 haloes.

## 2.7 Cold molecular gas and dark matter haloes

We saw in the previous sections that massive haloes were efficiently converting baryons into stars at high redshift. This seems to be in contradiction with Section 2.1 where we saw that gas cooling is more efficient at low redshift due to a higher metallicity of the gas. Moreover observations show that galaxies had on average a higher star formation rate at higher redshift, peaking at  $z \sim 2$  (Madau and Dickinson 2014), and that it can be linked to a higher cold molecular gas to baryonic mass fraction (Tacconi et al. 2010; Carilli and Walter 2013; Tacconi et al. 2018).

As the cold molecular gas is the basic ingredient of star formation, its presence in massive haloes at high redshift could explain their efficient star formation. In this section we test if evolution of the stellar mass to halo mass fraction can be linked to the evolution of the cold gas mass to stellar mass fraction. To do this, we use observations of the cold gas content of galaxies at high redshift and link the cold gas mass of central galaxies to the mass of their host dark matter halo.

### 2.7.1 Linking gas mass to stellar mass

Cold ( $T \sim 10$  K) and dense ( $n > 30 \text{ cm}^{-3}$ ) molecular clouds are mainly composed of  $\text{H}_2$  (Kennicutt and Evans 2012). However the  $\text{H}_2$  gas is almost invisible in emission as it is excited only for  $T \geq 100$  K. Fortunately the molecular interstellar medium also contains heavier elements at the level of a few  $10^{-4}$  per H nucleon, the most abundant being carbon and oxygen which combine in the CO molecule. The CO molecule is excited for  $T \geq 5$  K. As such the CO emission line is a good probe of the cold gas mass, dominated by  $\text{H}_2$  (see e.g. Bolatto et al. 2013). However this technique relies on spectroscopic observations and is therefore time consuming, especially to get a statistically significant sample at high redshift.

Recently, Scoville et al. (2016) proposed an empirical calibration of the Rayleigh-

Jeans dust emission continuum over the CO emission. They used local galaxies and galaxies at  $z \sim 2$  to calibrate this relation. This method gives a factor two accuracy on the derived gas mass, provided that we restrict the analysis to massive galaxies ( $> 10^{10} M_{\odot}$ ), as these are the ones where the gas is expected to follow a near-solar metallicity.

Using this calibrated relation, [Scoville et al. \(2017\)](#) were able to measure the gas mass ( $M_{\text{gas}}$ ) for 708 galaxies in the COSMOS field between  $0.3 < z < 4.5$ . These measurements were obtained with the Atacama Large Millimetre Array (ALMA) at 240 GHz and 345 GHz. Combining with observations of the COSMOS field in other wavelengths, they obtained the photometric redshift, the stellar mass and the star formation rates for the galaxies of their sample.

From this dataset they proposed an empirical relation between the gas mass of galaxies, their stellar mass  $M_*$ , their specific star formation rate sSFR and their redshift ([Scoville et al. 2017](#)). However, the equations provided in [Scoville et al. \(2017\)](#) are based on the star formation rate of main sequence<sup>6</sup> galaxies estimated in [Speagle et al. \(2014\)](#). Due to an error in the table giving the best fit parameters as a function of redshift in [Speagle et al. \(2014\)](#), the equations of [Scoville et al. \(2017\)](#) have been updated (N. Scoville, private communication). The gas mass, also called interstellar medium (ISM) mass, is given by

$$M_{\text{gas}} = 6.885 \times 10^9 \tau_{\text{evol}}^{0.5042} \left( \frac{\text{sSFR}}{\text{sSFR}_{\text{MS}}} \right)^{0.2802} M_{10}^{0.1206}, \quad (2.11)$$

where

$$\tau_{\text{evol}} \equiv \frac{\text{SFR}_{\text{MS}}(t, M_* = 5 \times 10^{10})}{\text{SFR}_{\text{MS}}(t_0, M_* = 5 \times 10^{10})}, \quad (2.12)$$

tracks the time evolution,  $M_{10} = M_*/10^{10} M_{\odot}$  and  $\text{sSFR} = \text{SFR}/M_*$  is the specific star formation rate. The  $\text{SFR}_{\text{MS}}$  is the SFR of main sequence galaxies and is taken from [Speagle et al. \(2014\)](#) and [Lee et al. \(2015\)](#).

We define the cold gas mass fraction as

$$f_{\text{gas}} \equiv \frac{M_{\text{gas}}}{M_* + M_{\text{gas}}}, \quad (2.13)$$

and fitted relation between the SFR and the cold gas mass is given by

$$\text{SFR} = 0.1052 \frac{M_{\text{gas}}}{10^9 M_{\odot}} \tau_{\text{evol}}^{0.5062} (\text{sSFR} / \text{sSFR}_{\text{MS}})^{0.7132} M_{10}^{0.2522} \quad (2.14)$$

## 2.7.2 Linking gas mass to halo mass

We now wish to link the above relation to the evolution of the dark matter halo mass to obtain the full picture of the evolution of the cold gas content contained in the central galaxies of dark matter haloes. In order to link the stellar mass in equation 2.11 to the dark matter halo mass, we use the stellar-to-halo mass relation

<sup>6</sup>The main sequence (MS) of galaxies is a tight relation followed by most galaxies which links their star formation rate to their stellar mass (e.g. [Noeske et al. 2007](#)).

published in Behroozi et al. (2018).

Behroozi et al. (2018) used a combination of low and high redshift surveys to constrain the SHMR and its evolution with redshift. Their stellar-to-halo mass relation is given by

$$\log_{10} \left( \frac{M_*}{M_1} \right) = \epsilon - \log (10^{-\alpha x} + 10^{-\beta x}) + \gamma \exp \left[ -0.5 \left( \frac{x}{\delta} \right)^2 \right], \quad (2.15)$$

$$x \equiv \log_{10} \left( \frac{M_h}{M_1} \right), \quad (2.16)$$

where each of the parameters ( $M_1$ ,  $\epsilon$ ,  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ) has a polynomial dependence on redshift (not reproduced here) given by equations J3 to J8 of Behroozi et al. (2018). Combining this equation with equation 2.11, we are able to track the gas content of central galaxies as a function of their halo mass and redshift.

### 2.7.3 Results

We show in Figures 2.15a to 2.15c the cold gas mass, star formation rate and cold gas mass fraction of main sequence galaxies as a function of their dark matter halo mass and redshift. The grey areas in the bottom of the Figures correspond to galaxies with a stellar mass lower than  $10^{10} M_\odot$ , which is the minimum mass of the calibrations of Scoville et al. (2017). The white dashed lines track the median mass accretion history of haloes, following the relations described in Behroozi et al. (2013) fitted on dark matter simulations (their Equations H2 to H6). These lines allow to track the previous mass of present day haloes. The top grey areas show haloes which will get a present mass higher than  $10^{15.5} M_\odot$  and are not expected to exist.

Figure 2.15a shows that the cold gas mass peaks at  $z > 2$  and for galaxies hosted in haloes more massive than  $M_h = 10^{12.5} M_\odot$ , reaching  $M_{\text{gas}} > 8 \times 10^9 M_\odot$ . Following the white dashed lines, we see that this cold gas will be depleted, and by  $z = 0.5$  the gas mass for all halo mass is lower than  $2 \times 10^9 M_\odot$ : it has decreased by a factor of four.

This depletion of cold gas is correlated with a decrease of the star formation rate, as we can see in Figure 2.15b. Star formation is higher at  $z > 2$  and is maximum in haloes with  $M_h > 10^{12.5} M_\odot$ , reaching more than  $30 M_\odot \text{yr}^{-1}$ . We see that the star formation rate decreases with time, going from  $\text{SFR} > 30 M_\odot \text{yr}^{-1}$  at  $z > 2$  to  $\text{SFR} < 5 M_\odot \text{yr}^{-1}$  for  $z < 1$ . These figures are coherent with observations of the decline of the SFR for main sequence galaxies with time (Noeske et al. 2007). Figure 2.15b, as well as 2.15a hint towards a dependence of the SFR with the halo mass at high redshift, while we see that SFR does not depend on mass at lower redshift.

Figure 2.15c shows the gas fraction of main sequence galaxies as a function of the host halo mass and redshift. We see in this Figure that for high redshift ( $z > 3$ ), more than 50 % of the baryonic content of central galaxies is in the form of gas. This is especially true for galaxies inside haloes of mass  $M_h < 10^{12.5} M_\odot$ , with a gas fraction of more than 60% at  $z > 2$ .

Following the white dashed lines we can track the growth of the haloes. Haloes of present day mass of  $10^{14} M_\odot$  had a mass of  $10^{12} M_\odot$  at  $z = 4$ . By  $z \sim 2$ , their

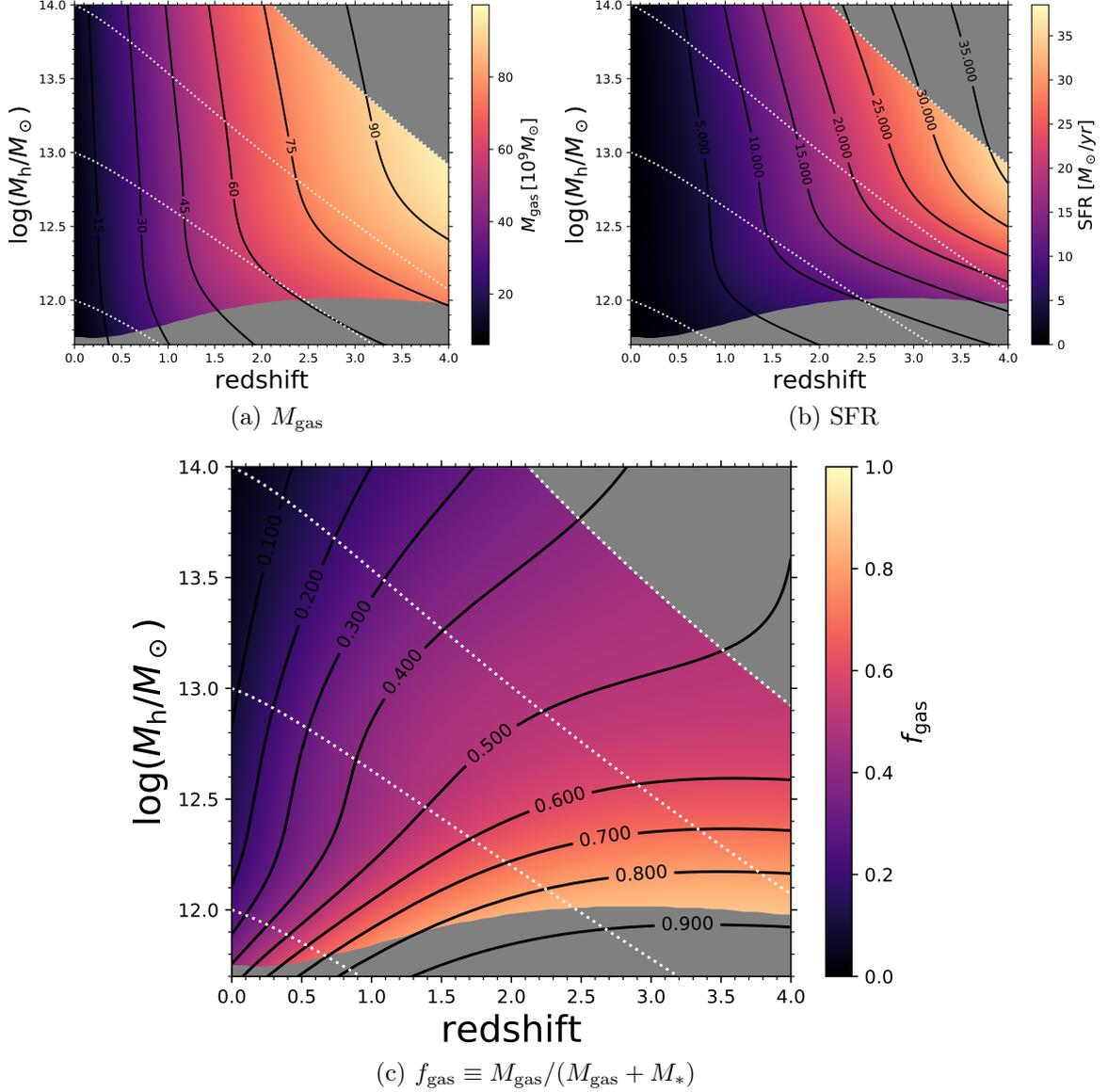


Figure 2.15: Properties of central galaxies as a function of their host halo mass and redshift, for main sequence galaxies ( $\text{sSFR} = \text{sSFR}_{\text{MS}}$ ). (a): Cold gas mass, (b): Star formation rate, (c): Cold gas fraction. Black lines show isocontours. The white dashed lines track the mass accretion history of haloes, and the upper right grey areas show location of haloes that are expected to have a present day mass of  $M_{\text{h}} > 10^{15.5} M_{\odot}$  and thus not expected to exist. The bottom grey areas show the locus of haloes hosting a galaxy with a stellar mass lower than  $M_* = 10^{10} M_{\odot}$ , results below this limit are extrapolated and should be taken with care.

central galaxies already converted 50% of their baryonic mass into stars. Central galaxies in haloes with a present day mass of  $10^{13} M_{\odot}$  have depleted 50 % of their gas at  $z \sim 1.5$  and the ones in haloes with a present day mass of  $10^{12} M_{\odot}$  have reached this threshold at  $z \sim 0.5$ .

Another way to track the evolution of gas is to track the halo mass and redshift of galaxies with a constant gas fraction, which are the black lines on [2.15c](#).

We see that the galaxies with a gas fraction of 60% have a halo mass of  $10^{12.5} M_{\odot}$  at  $z \sim 4$ . This mass stays almost constant until  $z \sim 2$ . And then we see that this mass quickly decreases down to  $\sim 10^{11.5} M_{\odot}$  at  $z = 0$ . This evolution of a constant gas fraction of 60% follows a trend similar to the evolution of the peak halo mass we observed in [Figure 2.10](#), with a plateau of the peak halo mass between  $z = 4$  and  $z = 2$  and then a decrease down to  $z = 0$ . This hints towards the importance of the depletion of gas on the evolution of the shape of the SHMR. We discuss more in [Section 2.8](#) the role of gas in the galaxy evolution and star formation quenching.

To summarize our results of the link between cold gas and halo mass across time:

- Galaxies with the highest gas mass and the highest star formation rate lie inside the most massive haloes at high redshift.
- However, the gas fraction is higher for galaxies inside a smaller halo.
- As halo mass grows, galaxies deplete their gas, decreasing both the gas fraction and the SFR with time.
- The gas mass fraction quickly decreases between  $z = 2$  and  $z = 0$  for haloes of a mass between  $10^{12}$  and  $10^{12.5} M_{\odot}$ .
- The redshift evolution of the halo mass for which galaxies have a gas fraction of 60 % is similar to the redshift evolution of the peak halo mass of the SHMR: an increase from  $z = 0$  to  $z = 2$ , then a plateau at  $M_h = 10^{12.5} M_{\odot}$  for  $z > 2$ .

## 2.8 What could explain a redshift evolution of quenching ?

Based on observations in the COSMOS field, we saw that the  $M_*/M_h$  ratio has a bell shape, with a peak around  $10^{12} M_{\odot}$  at  $z = 0$ . We interpreted this  $M_h^{\text{peak}}$  as the mass of haloes that have been most efficient in converting baryons into stars. Thanks to a hydrodynamical simulation, we found in [Section 2.6](#) that the star formation of galaxies is correlated with the stellar-to-halo mass relation. Star formation is very efficient in small haloes while haloes above  $M_h^{\text{peak}}$  have entered the quenching phase. As such,  $M_h^{\text{peak}}$  gives the characteristic mass of haloes that are entering the quenching phase.

We saw in [Figure 2.10](#) that in the COSMOS field,  $M_h^{\text{peak}}(z)$  increases with redshift, and changes slope at  $z \sim 2$ , showing a plateau at higher redshift. The efficiency of star formation has moved from massive haloes at high redshift, to less massive ones at low redshift. This implies that the threshold for massive haloes to enter the quenching phase depends on redshift: in the early universe quenching mechanisms are less effective for galaxies in haloes between  $10^{12}$  and  $10^{12.5} M_{\odot}$ .

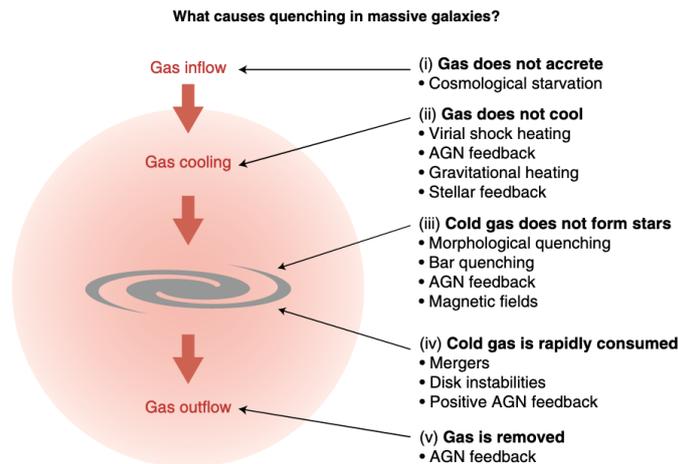


Figure 2.16: Figure from [Man and Belli \(2018\)](#). The role of gas and different physical processes of star formation quenching.

Cold molecular gas is the fuel of star formation. When the gas is accreted inside the dark matter halo, it first needs to cool, as we saw in Section 2.1. Several processes can halt galaxy formation, from external process that stop inflows of gas inside the halo, such as the merging into a bigger halo, to internal processes that prevent gas cooling such as AGN activity. Figure 2.16 lists some of the main physical processes that could explain quenching. In Section 2.7, we used observations of the cold molecular gas content of galaxies and we found that the cold molecular gas is mainly found at high redshift. We also found that the cold molecular gas mass fraction decreases drastically for haloes of a mass  $10^{12}$  to  $10^{12.5} M_{\odot}$  between  $z = 2$  and  $z = 0$ , similarly to the decrease of the peak halo mass from  $10^{12.5} M_{\odot}$  at  $z = 2$  to  $10^{12} M_{\odot}$  at  $z = 0$ .

We now need to find a scenario that can explain our observations of the evolution of  $M_h^{\text{peak}}$  with redshift, which we assume to be linked to our observations of the depletion of cold molecular gas in massive haloes.

In their cosmological hydrodynamical simulations, [Gabor and Davé \(2015\)](#) implement a heuristic prescription to halt star formation in systems with a large fraction of hot gas. Namely, their model prevents gas cooling in dark matter haloes by setting the circumgalactic gas temperature equal to the virial temperature. This condition is triggered when a halo has 60% of its gas particles with a temperature  $> 10^{5.4} \text{ K}$  ([Kereš et al. 2005](#)). The condition triggering the quenching phase, which [Gabor and Davé](#) call “hot halo” mode, is reached exclusively at  $M_h > 10^{12} M_{\odot}$  in their simulation. We find that this halo mass threshold is in agreement with  $M_h^{\text{peak}}$ . However, [Gabor and Davé \(2015\)](#) carried out their analysis at  $z < 2.5$ . For higher redshifts, this temperature threshold would be in disagreement with our results. Also [Behroozi et al. \(2018\)](#), considering the evolution of the quiescent galaxy fraction, emphasise that a quenching recipe with a constant temperature threshold could not explain the observational trend, see Figure 2.17. As the difference between a constant and a time-evolving threshold becomes more relevant in the first  $\sim 2 \text{ Gyr}$  after the Big Bang, our results could help discriminate between these different scenarios.

The hot halo model is agnostic regarding the sub-grid physics of the simulation:

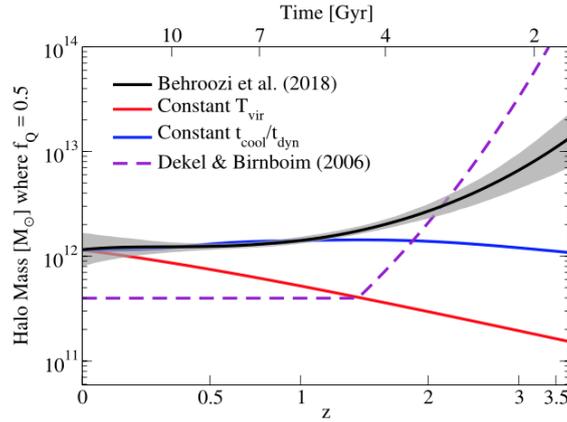


Figure 2.17: Figure from Behroozi et al. (2018). Mass of haloes above which 50% of the galaxy population is quenched, as a function of redshift. The plain red line assumes that quenching happens above a constant virial temperature threshold of  $6.4 \times 10^5$  K. The blue line assumes quenching happens when the ratio of the cooling time over the dynamical time is greater than 0.4. The purple dashed line is the model from Dekel and Birnboim (2006) assuming cold gas inflows penetrate the halo at  $z > 1.5$ . The black line are the results from the empirical model of Behroozi et al. (2018).

gas heating can be caused by other mechanisms (see Figure 2.16) such as stable virial shocks (Birnboim and Dekel 2003) or AGN feedback (see a review in Heckman and Best 2014). Simulations in Dekel and Birnboim (2006) show that shock heating in massive haloes becomes inefficient at high redshift because cold streams are still able to penetrate into the system and fuel star formation (see also Dekel et al. 2009). However, despite that this trend is in general agreement with our results there are quantitative differences in the evolutionary trend. With the fiducial parameters assumed in Dekel and Birnboim (2006) the “critical redshift” at which  $\sim 10^{12} M_\odot$  haloes start to form stars more efficiently is  $z_{\text{crit}} \simeq 1.5$ . Moreover, according to their model  $M_{\text{h}}^{\text{peak}}$  should keep increasing at  $z > z_{\text{crit}}$  instead of plateauing.

Quenching models more compatible with our observational results have been presented e.g. in Feldmann and Mayer (2015). Under the assumption that gas inflow (thus star formation) is strongly correlated to dark matter accretion, the authors note that at  $z > 2$  massive haloes are still collapsing fast and dark matter filaments efficiently funnel cool gas into the galaxy. At  $z \lesssim 2$  those haloes should enter in a phase of slower accretion that eventually impedes star formation by gas starvation. However, we caution that they study single galaxies in cosmological zoom-in simulations: a larger sample may show considerable dispersion in the redshift marking the transition between the two dark matter accretion phases. In addition, we emphasise that not only the accretion rate but also the cooling timescale varies with redshift. Gas density follows the overall matter density of the universe, evolving as  $\propto (1+z)^3$ . Since the post-shock cooling time is proportional to gas density, it would be significantly shorter at higher redshift. On the other hand, this argument in absence of more complex factors should lead to a steeper, monotonic increase of  $M_{\text{h}}^{\text{peak}}$  that we do not observe.

As mentioned above, AGN feedback at high redshifts can also regulate galaxy star formation and explain our observed  $M_h^{\text{peak}}$  trend. AGN activity at high redshift is expected to be almost exclusively in quasar mode (e.g., [Silk and Rees 1998](#)) with powerful outflows that can heat or even expel gas. However such radiative feedback has shown to be inefficient in hydrodynamical “zoom-in” simulations at  $z \sim 6$  (e.g. [Costa et al. 2014](#)). Observations also indicate that high- $z$  quasars do not prevent significant reservoirs of cold gas from fuelling star formation (e.g., [Maiolino et al. 2012](#); [Cicone et al. 2014](#)). Therefore, star formation in massive haloes can proceed for 2–3 Gyr after the Big Bang without being significantly affected by AGN activity, in agreement with our observations. At later times, perturbations to cold filamentary accretion can starve galaxies of their gas supplies ([Dubois et al. 2013](#)).

Models of galaxy formation efficiency discussed so far do not account for the role of large scale structures. Depending on their location within the cosmic web (filaments, nodes, voids) haloes with similar masses may experience different accretion histories ([De Lucia et al. 2012](#)). One key idea in this context is “cosmic web detachment” ([Aragon-Calvo et al. 2016](#)): galaxies tied to nodes or filaments are removed from their original location by interaction with another galaxy. After the detachment gas supply – and then star formation – becomes less efficient. [Aragon-Calvo et al. \(2016\)](#) suggest that massive haloes are the first to detach, whereas less massive haloes  $0.1\text{--}3 \times 10^{10} h^{-1} M_\odot$  are still part of the cosmic web today. It is difficult to test this scenario beyond the local universe because precise measurements of the SMF are required in addition to higher-order statistics (e.g., 3-point correlation functions).

In summary, we have described different physical processes which could explain our observed trends. The complete answer is likely to be some combination of these mechanisms. But based on our discussion and observations, the physical processes at work in our results seem to be best understood as a combination of starvation of the cold-flow accretion and AGN feedback, with the precise role of evolutionary and environmental effects yet to be determined.

## 2.9 Conclusions

We have used a sub-halo abundance matching technique combined with precise stellar mass function measurements in COSMOS to measure the stellar-to-halo mass relation with one coherent sample, on an unprecedented redshift interval from  $z \sim 0$  to  $z \sim 5$ . We accounted for the main sources of uncertainties in our stellar mass measurements and photometric redshifts. We also tested the impact of halo mass function uncertainties on the resulting SHMR. At  $z \sim 0.2$  we found that the ratio of mass in stars to dark matter halo mass ( $M_*/M_h$ ) peaks at a halo mass of  $10^{12.05 \pm 0.07} M_\odot$ . This peak mass increases steadily to  $10^{12.48 \pm 0.08} M_\odot$  at  $z \sim 2.3$ , and remains almost constant up to  $z = 4$ .

In order to validate our findings on the evolution of the peak halo mass, we have used the lightcone catalog of the Horizon-AGN simulation, where galaxies and haloes have been linked. We obtained a global view of the impact of the stellar-to-halo mass relation on star formation rate of galaxies. Using a qualitative analysis, we found that haloes with a mass higher than  $M_h^{\text{peak}}$  are typically quenched, while

small haloes and hosting a small galaxies are very efficient in forming stars.

We have used observations of the cold gas content in galaxies at high redshift in the COSMOS field to follow the cold molecular gas and the efficiency of galaxy formation. We found that the gas content is mainly contained in massive haloes at high redshift. However, we found that haloes of  $10^{12} M_{\odot}$  host a central galaxy with a gas fraction higher than 80% at  $z > 1.7$ . Following the evolution of halo growth, we found that the redshift at which the central galaxy reaches a gas fraction of 50 % depends on the initial mass of the halo. Following a line of constant gas fraction, we found similarity with the evolution of the peak halo mass, hinting to a link between gas depletion and quenching of star formation.

We discussed qualitatively which physical processes control these observations. We propose that these evolutions can either be related to AGN feedback or to environmental effects such as the starvation of cold gas inflows at high redshift, maybe due to cosmic web detachment.

Our studies are based on phenomenological models and as a result can provide no direct information concerning the physical processes acting inside haloes. Next-generation hydrodynamical simulations will allow us to better understand the small-scale physical processes acting inside dark matter haloes and determine what physical effects control star formation. In the next few years, the combined 20 deg<sup>2</sup> Spitzer–Euclid legacy and Hawaii-2-0 surveys on the Euclid deep fields will provide much better constraints on the massive end of the SMF at high redshifts. Precise photometric redshifts will allow us to investigate in detail the role of environment and in particular the “cosmic web” role in shaping galaxy and dark matter evolution.

Table 2.1: Best fit parameters for the ten redshift bins with their 68 per cent confidence intervals, and  $M_{\text{H}}^{\text{peak}}$  recovered from the best fit SHMR with its 68 per cent confidence interval.

Redshift bin	$\log(M_1/M_{\odot})$	$\log(M_{*,0}/M_{\odot})$	$\beta$	$\delta$	$\gamma$	$\xi$	$\log(M_{\text{H}}^{\text{peak}}/M_{\odot})$
[0.2, 0.5]	$12.49^{+0.13}_{-0.094}$	$10.84^{+0.11}_{-0.077}$	$0.463^{+0.040}_{-0.030}$	$0.77^{+0.16}_{-0.29}$	$< 0.802$	$0.138^{+0.034}_{-0.066}$	$12.05 \pm 0.07$
[0.5, 0.8]	$12.668^{+0.089}_{-0.074}$	$11.039^{+0.074}_{-0.060}$	$0.458^{+0.026}_{-0.023}$	$0.81^{+0.17}_{-0.24}$	$< 0.723$	$0.099^{+0.022}_{-0.027}$	$12.22 \pm 0.05$
[0.8, 1.1]	$12.614^{+0.073}_{-0.060}$	$11.006^{+0.056}_{-0.042}$	$0.437^{+0.025}_{-0.022}$	$0.93^{+0.19}_{-0.28}$	$< 0.955$	$0.088 \pm 0.015$	$12.14 \pm 0.04$
[1.1, 1.5]	$12.642^{+0.086}_{-0.069}$	$10.978^{+0.072}_{-0.054}$	$0.407^{+0.029}_{-0.023}$	$0.80^{+0.16}_{-0.23}$	$< 0.629$	$0.092^{+0.023}_{-0.025}$	$12.26 \pm 0.04$
[1.5, 2.0]	$12.78^{+0.10}_{-0.072}$	$11.053^{+0.080}_{-0.055}$	$0.438^{+0.035}_{-0.026}$	$0.82^{+0.17}_{-0.25}$	$< 0.724$	$0.075 \pm 0.017$	$12.35 \pm 0.04$
[2.0, 2.5]	$13.062^{+0.078}_{-0.087}$	$11.15^{+0.11}_{-0.095}$	$0.525^{+0.033}_{-0.027}$	$1.09^{+0.36}_{-0.68}$	$< 2.08$	$0.128^{+0.045}_{-0.050}$	$12.48 \pm 0.08$
[2.5, 3.0]	$13.11 \pm 0.18$	$11.09 \pm 0.25$	$0.598^{+0.045}_{-0.036}$	$1.01^{+0.55}_{-0.72}$	— —	$0.216^{+0.061}_{-0.14}$	$12.47 \pm 0.19$
[3.0, 3.5]	$13.14^{+0.22}_{-0.20}$	$11.14 \pm 0.27$	$0.631^{+0.071}_{-0.038}$	$0.73^{+0.35}_{-0.54}$	$< 2.47$	$0.176^{+0.074}_{-0.085}$	$12.49 \pm 0.17$
[3.5, 4.5]	$13.30^{+0.20}_{-0.27}$	$11.41^{+0.28}_{-0.46}$	$0.625^{+0.056}_{-0.039}$	— —	$< 2.93$	$0.231 \pm 0.099$	$12.63 \pm 0.25$
[4.5, 5.5]	$14.35^{+0.89}_{-1.0}$	$< 13.5$	$0.642^{+0.094}_{-0.11}$	— —	— —	$0.45^{+0.22}_{-0.34}$	$13.35 \pm 0.54$

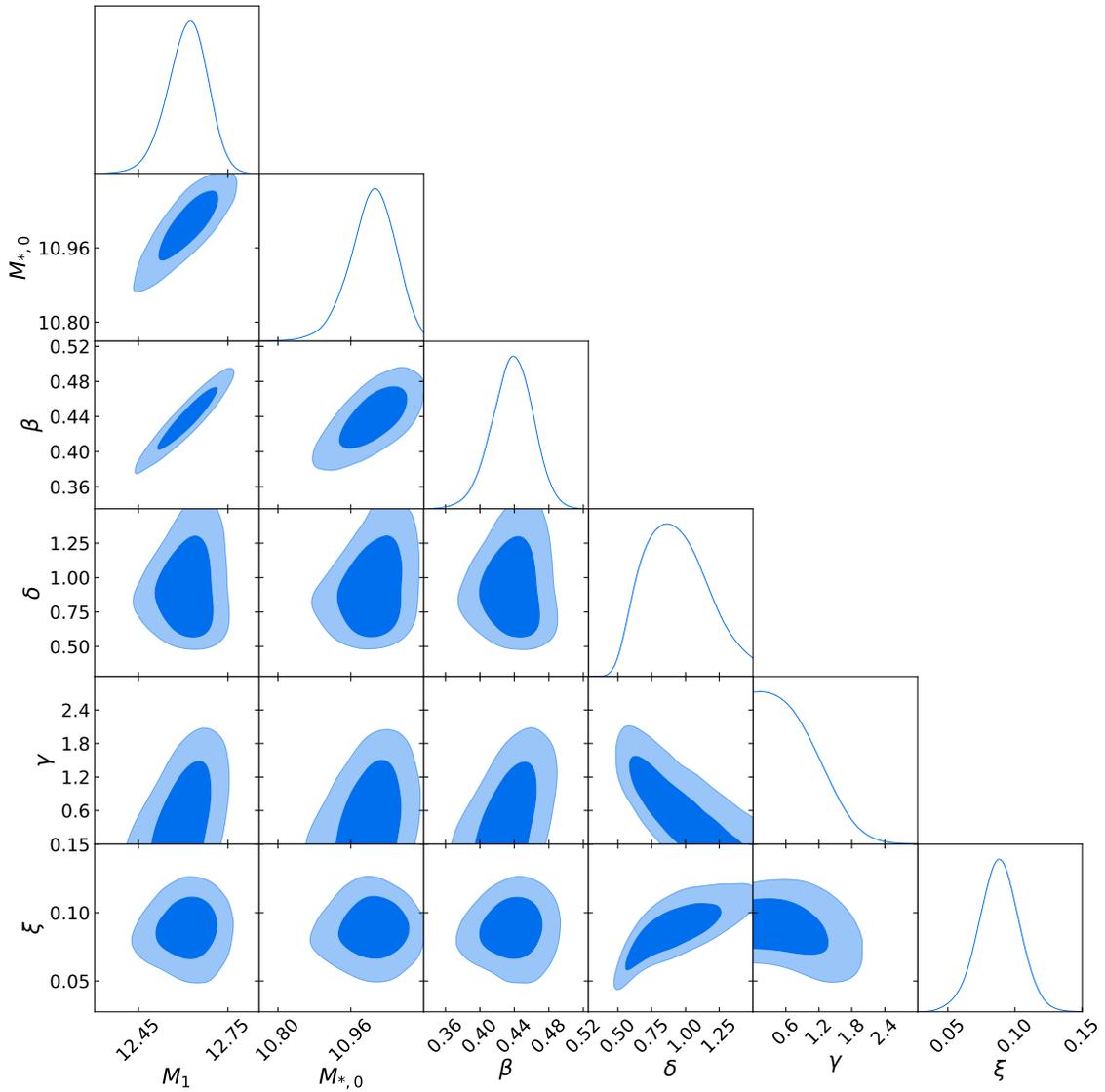


Figure 2.18: Figure from [Legrand et al. \(2019\)](#). One and two dimensional marginalised distributions for the six free parameters of our SHMR (Equation 2.6) in the redshift bin  $[0.8, 1.1]$ . Solid contours give the 68 and 95 per cent confidence intervals.

## Chapter 3

# The cosmological power of a joint analysis of *Euclid* and CMB surveys

The last twenty years saw the triumph of the standard cosmological model. With the  $\Lambda$ CDM model, we were able to model with very high fidelity the power spectrum of the CMB observed by the *Planck* experiment. Concurrently, the galaxy clustering and the BAO signals observed by galaxy surveys such as SDSS are also perfectly compatible with  $\Lambda$ CDM. However, having a model that perfectly fit our observations does not mean that we fully understand our Universe. In reality, the  $\Lambda$ CDM model does not provide any clue about the nature of the dark matter nor of the dark energy. Understanding the nature of this dark sector is the key question for the next decades in cosmology.

To investigate the dark sector, next generation galaxy surveys will map the distribution of matter with unprecedented precision, offering a window on the evolution of the Universe. Typically, the *Euclid* survey will probe a large volume of the universe, on 70% of the sky and up to a redshift of  $z = 2.5$ . This will provide a census of billions of galaxies with a photometric redshift estimation, and millions of them with a spectroscopic redshift. By targeting the era where the dark energy started to dominate the energy budget of the Universe, we will surely discover a lot of information about the nature of this accelerated expansion.

In parallel, next generation CMB surveys will map the microwave sky at a few arcminutes resolution, both in temperature and polarisation. These surveys will allow us to get a precise view of the early Universe. Moreover, because the CMB photons crossed all the Universe up to us, we will get an integrated view of the distribution of matter in the Universe (mainly thanks to CMB lensing) and on how it has evolved.

This arrival of new and exquisite datasets requires that we develop the tools to retrieve optimally the cosmological information they contain. Indeed, the era of precision cosmology has now arrived: we are not limited by statistical uncertainties anymore, our limits will be the systematic uncertainties. Because of the intrinsic different nature of the CMB and galaxy probes, they are affected by different systematics. Combining and cross-correlating these datasets will be key to overcome the limitations due to the systematic uncertainties and to get the tightest constraints on our cosmological model.

One necessary step is to assess how well these experiments will perform in terms of constraints on cosmological parameters. These forecasts are key to provide the reference needed to check the performance of the surveys and validate the choices made during their design. These forecasts will also provide the templates and reference tools for the future data analysis.

In this context, I will detail in the present chapter forecasts on the combination of *Euclid* and CMB observables. I will first introduce in Section 3.1 the interest of probe combinations, before presenting our forecasting formalism in Section 3.2. I will then introduce the *Euclid* survey in Section 3.4 and our observables of interest in Section 3.3. I will then detail the numerical computations in Section 3.6. Finally I will present our results in Section 3.7, focusing on the performance gain we will have when including CMB observables to the main *Euclid* analysis, and I will conclude in Section 3.8.

This chapter introduces the work of the *Euclid* XCMB science working group, dedicated to the cross correlations of *Euclid* and CMB data. I present the work of the SWG and my contribution to the activities of the group and to the article that will be submitted by the collaboration ([Euclid Collaboration in prep.](#)). More specifically, my participation in this project was dedicated to the modelling of CMB noise power spectra, to the implementation and production of the covariance matrices between our probes, of the Fisher matrices, and to the writing of the paper. Moreover, by checking for consistency in the Final Fisher matrices between the different codes I participated in the comparison and inter-validation of the codes developed in our analysis.

In this chapter and in the next one, we use the following naming conventions: *observable* refers to a field built on measured quantities, such as counts, redshifts, or deflection angles, while *probe* refers to the combination of one, two or more observables into a summary statistics. In practice, our probes will be the two-point angular power spectra  $C_\ell$ , and the two-point 3D power spectrum  $P(k)$ . When combining two different observables we will talk about a cross-correlation, while the correlation of the same observable is an auto-correlation.

## 3.1 Interests of probe combination

The interest of probes combination has been widely demonstrated (see e.g. [Eisenstein et al. 1999](#); [Lewis and Bridle 2002](#)). Combining probes help in breaking degeneracies between cosmological parameters, in lowering the importance of systematic effects and in checking for inconsistencies between the different datasets.

In a recent example, the Figure 3.1 from [Suzuki et al. \(2012\)](#) shows that combining supernovae observations with CMB and BAO constraints allows to break degeneracies between  $\Omega_\Lambda$  and  $\Omega_m$ , and show that the Universe is almost completely flat. However, in this paper they assumed that the observables were uncorrelated, which is valid due to the very different nature of each of them: CMB is an early probe while BAO measurements was done at  $z < 1$  ([Percival et al. 2010](#)) and supernovae can be considered as a local probe uncorrelated with the others.

In future surveys, we will not be able to assume a total independence between CMB and galaxy probes, and a careful modelling of their covariances is needed to

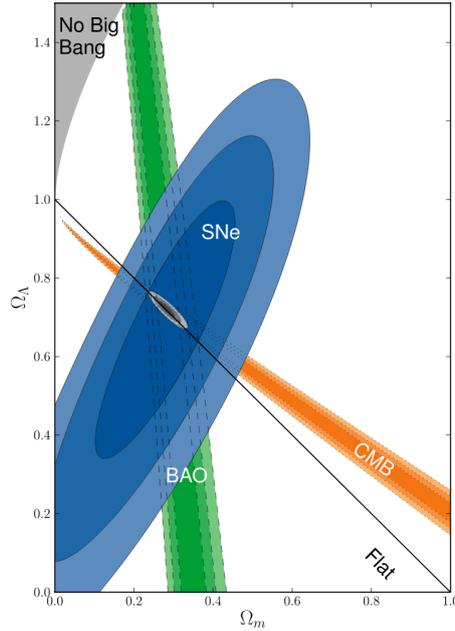


Figure 3.1: Figure from [Suzuki et al. \(2012\)](#) showing the degeneracy between  $\Omega_\Lambda$  and  $\Omega_m$  from different datasets: the CMB constraints from WMAP in orange ([Komatsu et al. 2011](#)), the BAO constraints from SDSS in green ([Percival et al. 2010](#)) and the Supernovae constraints from [Suzuki et al. \(2012\)](#) in blue. The grey area shows the constraints obtained when combining the three.

perform accurate joint analyses. This is especially true when the footprints of the CMB and galaxy surveys overlap, i.e. for *Euclid* and *Planck*, or for the Dark Energy Survey (DES) and the South Pole Telescope (SPT). Indeed the matter density field traced by the galaxies is correlated with the CMB observables, as the interaction of CMB photons with the large scale structures will create secondary anisotropies (see Section 1.2.2).

It is common practice to perform a joint analysis between different observables from the same experiment. For instance the [Planck Collaboration et al. \(2018a\)](#) analysis used the cross-power spectra between the temperature, polarisation and CMB lensing potential fields to get the final constraints on cosmological parameters. In this case, because all observables are taken from the same observation, it is easy to cross-correlate them and to correctly take into account their covariances.

Similarly, weak lensing galaxy surveys like DES are using the auto- and cross-correlations between galaxy weak lensing and galaxy clustering to constrain cosmological parameters ([Abbott et al. 2018](#)). In this case also, each observable comes from the same dataset so they can easily be cross-correlated.

Performing a joint analysis between two completely different datasets is much more difficult. However, it has been shown to be a very promising way to break degeneracies between cosmological and astrophysical systematic parameters. Indeed [Vallinotto \(2013\)](#), [Pearson and Zahn \(2014\)](#), [Merkel and Schäfer \(2017\)](#) and [Schaan et al. \(2017\)](#) have shown that a joint analysis of CMB lensing and galaxy probes for next generation surveys will help in breaking degeneracies between the amplitude of the matter power spectrum, the galaxy bias and the weak lensing calibration

bias. These degeneracies breaking with systematic parameters such as the galaxy bias will greatly improve the constraints on cosmological parameters, such as the sum of neutrino masses as shown in [Giusarma et al. \(2018\)](#). Moreover, the cross correlations of galaxy surveys with iSW and CMB lensing was shown to be useful to constrain primordial non-Gaussianity ([Giannantonio et al. 2014](#); [Giannantonio and Percival 2014](#)).

Joint analysis are now planned early on in the development stage. Indeed, the Dark Energy Survey and the South Pole Telescope have been designed to maximize the overlap between the galaxy and the CMB observations. Doing so they are able to cross-correlate galaxy observables with the CMB lensing field. They chose to perform a joint analysis by cross-correlating the galaxy weak lensing, the galaxy clustering and the CMB lensing fields ([Omori et al. 2017](#); [Abbott et al. 2019](#)) This combination of probes helped in checking for inconsistencies between the datasets and was a powerful confirmation test for the results of each experiment.

The above examples mainly concern cross-correlation of the CMB lensing field with galaxy observables. Another possible cross-correlation is between CMB temperature and CMB polarization with galaxy observables. Most notably, on large scales ( $\ell < 30$ ) the cross correlation of CMB temperature with galaxy clustering will be dominated by the late-time iSW effect which is due to the accelerated expansion (see e.g. [Douspis et al. 2008](#); [Dupé et al. 2011](#); [Nishizawa 2014](#)). Even if the iSW has a very low signal to noise ratio (lower than 4), this effect is expected to be very sensitive to dark energy and modified gravity models. See e.g. [Challinor and Lewis \(2011\)](#) for details on the physical effects probed by the cross correlation of galaxy density with CMB temperature and polarisation.

## 3.2 Cosmological forecasts

Our scope is to forecast the improvements on cosmological parameters when performing a joint analysis of *Euclid* main observables with CMB observables, such as temperature, polarisation and CMB lensing potential. In order to assess the performances of upcoming experiments, we will use the Fisher formalism. We first introduce below the principles of cosmological analysis with a likelihood function, before describing the derivation of the Fisher matrix.

### 3.2.1 The likelihood function

The likelihood function allows to test the validity of a model against data taken from observations. More precisely, this function is defined as the probability that a given experiment would get the data it did for a given theory

$$\mathcal{L}(\mathbf{d}; \boldsymbol{\theta}) \equiv P(\mathbf{d}|\boldsymbol{\theta}) , \quad (3.1)$$

where  $\mathbf{d}$  is the data vector obtained from the experiment and  $\boldsymbol{\theta}$  is the set of parameters that describes our model.

We are interested in evaluating the parameters of the model that best reproduce the data,  $P(\boldsymbol{\theta}|\mathbf{d})$ . Thanks to the Bayes theorem, this probability is linked to the

likelihood function with

$$P(\boldsymbol{\theta}|\mathbf{d}) = \frac{P(\boldsymbol{\theta})P(\mathbf{d}|\boldsymbol{\theta})}{P(\mathbf{d})} = \frac{P(\boldsymbol{\theta})\mathcal{L}(\mathbf{d};\boldsymbol{\theta})}{P(\mathbf{d})}. \quad (3.2)$$

$P(\boldsymbol{\theta}|\mathbf{d})$  is called the *posterior*,  $P(\boldsymbol{\theta})$  is the *prior* and  $P(\mathbf{d})$  is a normalisation factor.

The posterior is what we want to measure. The maximum of this distribution is the most probable value of  $\boldsymbol{\theta}$ , and the width of the distribution gives the confidence interval we have on this estimation.

The prior represents the information we have before performing the experiment. It could come from a previous experiment or from theory. A flat prior (i.e. a constant  $P(\boldsymbol{\theta})$ ) assumes that all values are equiprobable. In this case, we can directly link the posterior to the likelihood

$$P(\boldsymbol{\theta}|\mathbf{d}) \propto \mathcal{L}(\mathbf{d};\boldsymbol{\theta}). \quad (3.3)$$

The parameters maximising the likelihood will be noted  $\boldsymbol{\theta}_{\text{ML}}$ . It is a function of the data vector  $\mathbf{d}$ . Finding the maximum of the likelihood is the key question once the likelihood is defined. It can be achieved through several possible maximum likelihood estimators, or by sampling the parameter space with a Monte-Carlo algorithm.

If we assume that the data vector is a random realisation taken from a multivariate Gaussian distribution, defined with a covariance matrix  $\mathbf{C}$ , the likelihood function is

$$\mathcal{L}(\mathbf{d};\boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2}\sqrt{\det \mathbf{C}}} \exp\left(-\frac{1}{2}(\mathbf{d} - \mathbf{d}^{\text{th}}(\boldsymbol{\theta}))^T \mathbf{C}^{-1}(\mathbf{d} - \mathbf{d}^{\text{th}}(\boldsymbol{\theta}))\right), \quad (3.4)$$

where  $N$  is the length of the data vector,  $\mathbf{d}^{\text{th}}$  is the theoretical data vector obtained with parameters  $\boldsymbol{\theta}$  and  $T$  denotes the transposition operation.

In general the covariance matrix depends on the parameter  $\boldsymbol{\theta}$ . However it is computationally costly to get the covariance matrix for each value of the parameter space when we are trying to find the maximum of the likelihood. This is why it is often assumed that the covariance matrix is independent of the parameters. This common approximation may lead to overestimations of the parameter constraints, especially when we start to be limited by systematic uncertainties (see e.g. [Eifler et al. 2009](#); [Krause and Eifler 2017](#)).

One of the difficult points of a likelihood analysis is to estimate the covariance matrix. There are three main methods to estimate the covariance matrix: from the data themselves, from simulations or from the theory.

If we estimate the covariance matrix from the data, we then need a large number of data realisations to correctly sample all the possible values of the data. Assuming that we have  $n$  realisations with  $\mathbf{d}^{(k)}$  one of the  $n^{\text{th}}$  realisations, an estimate of the covariance matrix is

$$\hat{\mathbf{C}} = \frac{1}{n-1} \sum_{k=1}^n (\mathbf{d}^{(k)} - \boldsymbol{\mu})(\mathbf{d}^{(k)} - \boldsymbol{\mu})^T, \quad (3.5)$$

where  $\boldsymbol{\mu}$  is the average of the data realisations ([Hartlap et al. 2007](#); [Dodelson and](#)

Schneider 2013; Taylor et al. 2013). If the average  $\boldsymbol{\mu}$  is not estimated from the data but is known in advance, then the denominator is  $n$  not  $n - 1$ .

The problem in cosmology is that we cannot get several realisations of our Universe, as opposed to experimental physicists who could repeat their experiments several times. For instance, we can only observe one image of the CMB, among the many possible realisations based on the same cosmological model. To overcome this we can rely on simulations (Blot et al. 2015; Schneider et al. 2020). The scope is then to produce several realisations of the data to get as close as possible to the actual data taken by the experiment. The covariance matrix is then estimated with Equation 3.5 with the data vectors taken from the simulations. The inherent limitations of numerical simulations are then the computational power needed to get both a large volume and a small resolution. This is even more difficult when we have to estimate the covariance matrix for each values of the parameter space.

Finally, one could rely on a theoretical estimation of the covariance matrix. In this case, the covariance includes two contributions, one which is the fiducial covariance matrix assuming the data vectors follow perfectly the theory, and one which models the noise due to the experiment. The limitations here are that we must accurately model all the covariance terms, especially the non-Gaussian terms due to the coupling between different Fourier modes created by non-linear gravitational interactions, and the super-sample covariance, due to the effect of a large scale background density mode in the survey window (see e.g. Krause and Eifler 2017; Sellentin and Heavens 2018; Lacasa 2018). In the present study we use a theoretical estimation of the covariance matrix and we neglect all non-Gaussian terms, following Euclid Collaboration et al. (2019). In future works, especially when the data will be analysed, a more realistic modelling of the covariance matrix will have to be implemented.

### 3.2.2 Fisher analysis

Fisher analysis allows to forecast of how well a given experiment will constrain parameters of a given model before actually taking any data or running any simulation. It is very useful to design experiments and check there performances. Its first use in a cosmology analysis can be dated to Tegmark (1997), and Tegmark et al. (1997) for both CMB and galaxy clustering analysis.

A straightforward way to obtain the Fisher matrix is to expand the likelihood in Taylor series close to its maximum

$$\ln \mathcal{L}(\boldsymbol{\theta}) = \ln \mathcal{L}(\boldsymbol{\theta}_{\text{ML}}) + \frac{1}{2} \sum_{\alpha\beta} (\theta_\alpha - \theta_{\alpha,\text{ML}}) \left. \frac{\partial^2 \ln \mathcal{L}}{\partial \theta_\alpha \partial \theta_\beta} \right|_{\text{ML}} (\theta_\beta - \theta_{\beta,\text{ML}}), \quad (3.6)$$

where the derivatives are computed at the location of the maximum of the likelihood, and the first derivative of the likelihood is zero at the location of the maximum by definition. As we are simulating the data and the experiment, the maximum of the likelihood is known by definition, we do not need to define a maximum likelihood estimator. As we see here, close to its maximum the likelihood can always be approximated by a multivariate Gaussian in the parameter space.

The Fisher matrix is defined as the ensemble average (over the data) of the Hessian matrix of Equation 3.6 (see e.g. Verde 2010)

$$F_{\alpha\beta} \equiv - \left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial \theta_\alpha \partial \theta_\beta} \right\rangle_{\boldsymbol{\theta}_{\text{ML}}} , \quad (3.7)$$

where the average here is the ensemble average, which averages over all the possible realisations of our Universe given the cosmological model defined by  $\boldsymbol{\theta}_{\text{ML}}$ .

In fact, as can be seen in equation 3.6, the Fisher matrix corresponds to the inverse of the covariance matrix of a Gaussian likelihood in the parameter space.

$$\mathcal{C} = F^{-1} \quad (3.8)$$

If the Fisher matrix is diagonal, then the covariance matrix  $\mathcal{C}$  is diagonal and the parameters are uncorrelated. Note that we previously have shown a Gaussian covariance between the data vectors (with covariance  $\mathcal{C}$ ), while here we have a Gaussian covariance between the parameters (with covariance  $\mathcal{C}$ ), we do not need to make the assumption of a Gaussian covariance of the data to get the Fisher matrix.

The expected uncertainties on each parameter are then given by

$$\sigma_\alpha = \sqrt{(F^{-1})_{\alpha,\alpha}} . \quad (3.9)$$

Here we have performed a *marginalisation* on the other parameters, i.e. we let them vary freely. If we assume that all other parameters are fixed, then the uncertainty on the parameter  $\theta_\alpha$  would be  $1/\sqrt{F_{\alpha,\alpha}}$ .

Marginalising on all but two parameters is done by selecting the rows and columns of two parameters of interest in the covariance matrix  $\mathcal{C}_{\alpha,\beta}$ . We thus obtain a  $2 \times 2$  covariance matrix which allow to easily plot the two dimensional confidence region between the parameters. As we assumed a Gaussian covariance since the beginning, the confidence regions drawn from a Fisher analysis will always be ellipses.

One commonly used quantity is the figure of merit (FoM) between two parameters. It is proportional to the area of the ellipse and is given by

$$\text{FoM} = \sqrt{\det F_2} , \quad (3.10)$$

where  $F_2$  is the marginalised  $2 \times 2$  Fisher matrix, taken as the inverse of the  $2 \times 2$  covariance matrix where we selected the rows and columns corresponding to parameters  $\alpha$  and  $\beta$

$$F_2 = (\mathcal{C}_{\alpha,\beta})^{-1} . \quad (3.11)$$

### 3.2.3 MCMC as an alternative to Fisher

Among the alternatives to the Fisher matrix analysis for forecasting experiments, the Markov Chain Monte Carlo (MCMC) analysis is often used. This method samples the parameter space with a Monte Carlo algorithm in order to get an estimation of maximum likelihood and to get the size of the confidence intervals.

The advantage of an MCMC analysis is that it is able to sample non-Gaussian posterior distributions. The Fisher analysis hypothesis that the likelihood is well approximated by a Gaussian in the parameter space can be completely wrong in some cases. As such the Fisher matrix could underestimate the errors on the parameters (see e.g. [Wolz et al. 2012](#); [Sellentin et al. 2014](#)).

However, the limit of a MCMC analysis is that it needs a lot of computation of the likelihood to efficiently sample the parameter space. This can be computationally costly especially when there is a large number of parameters. Fisher analysis is a much faster alternative which allows to test a lot of different scenarios.

### 3.2.4 Fisher matrix using angular power spectra

We will use the angular power spectra as our data vectors. In this case, one can show (see e.g. [Verde 2010](#)) that the Fisher matrix can be expressed as

$$F_{\alpha\beta} = \frac{\partial C_\ell}{\partial \theta_\alpha} \mathbf{C}^{-1} \frac{\partial C_\ell}{\partial \theta_\beta}, \quad (3.12)$$

where  $\mathbf{C}$  is the covariance matrix between the  $C_\ell$ .

This definition of the Fisher matrix for the  $C_\ell$  is often obtained in the literature by assuming that the  $C_\ell$  follow a Gaussian statistics. In practice, the  $C_\ell$  are not Gaussian, they follow a Wishart distribution, but for high  $\ell$  this can be approximated as a Gaussian. However, the  $a_{\ell,m}$  from which the  $C_\ell$  are derived, are Gaussian. So if one uses directly the  $a_{\ell,m}$  as data vector, the likelihood is Gaussian in the data space and Equation 3.4 applies, so we can easily derive the Fisher matrix using Equation 3.7. One can show that deriving the Fisher matrix when using the  $a_{\ell,m}$  as a data vector gives the same results as deriving the Fisher matrix with  $C_\ell$  as a data vector and by correctly defining their likelihood with a Wishart distribution ([Carron 2013](#)). As such, Equation 3.12 does not need to make the assumption that the  $C_\ell$ 's are Gaussian, it is the exact derivation of the Fisher matrix when using the true Wishart distribution of the  $C_\ell$ 's.

## 3.3 Modelling observables

The *Euclid* main observables are the spectroscopic galaxy density field (GCs), the weak lensing (WL) tomography and the photometric galaxy density (GCp) tomography. The 3D galaxy clustering probe is performed with the spectroscopic galaxy survey, while the WL and GCp fields from the photometric survey are used to compute 2D angular power spectra (auto and cross). The CMB observables we are considering are the CMB temperature ( $T$ ), the CMB polarisation ( $E$  modes) and the CMB lensing potential ( $\phi$ ). Given that our fiducial model does not include any tensor modes, we ignore the  $B$  modes from polarisation.

Our analysis includes all the auto- and cross spectra between the 2D observables of *Euclid* and of the CMB, namely GCp, WL,  $T$ ,  $E$  and  $\phi$ . However, because the spectroscopic galaxy clustering is a 3D probe, and due to the difficulty of correctly taking into account covariances between 3D and 2D probes, we assume that the GCs

probe is not correlated to the others. In practice in our Fisher analysis, we will sum the Fisher matrix of the 3D galaxy clustering to the Fisher matrix of our combined analysis of the 2D probes. Taking into account the correlation between a 2D and 3D probe is not trivial. Some approaches are being developed, such as [Passaglia et al. \(2017\)](#) and [Camera et al. \(2018\)](#), but they imply to replace the traditional 3D galaxy clustering probe with a new formalism more easily compatible with 2D probes.

We develop below the formalism to compute the auto- and cross power spectra of our 2D observables. The general expression to model the angular power spectra between two observables A and B is given by

$$C_\ell^{AB} = \frac{2}{\pi} \int_{z_1} \frac{c dz_1}{H(z_1)} \int_{z_2} \frac{c dz_2}{H(z_2)} \int_k k^2 dk W_A(k, z_1) W_B(k, z_2) \times j_\ell(k r(z_1)) j_\ell(k r(z_2)) P(k|z_1, z_2), \quad (3.13)$$

where the  $P(k|z_1, z_2)$  is the 3D matter power spectrum between two redshifts,  $j_\ell$  are the spherical Bessel functions of order  $\ell$ , and  $W(k, z)$  are the weight functions of the observables A and B. The weight functions  $W(k, z)$  contain all the information about the observables. They relate the underlying matter power spectrum to the observables. See e.g [Seljak and Zaldarriaga \(1996\)](#) for CMB observables, [Huterer et al. \(2001\)](#) for galaxy density observables and [Kilbinger \(2015\)](#) for the weak lensing observable.

Measurements by *Euclid* will extend down to small scales and we must take into account the non-linear evolution of the power spectrum. The matter power spectrum is computed with the non-linear corrections of *Halofit* [Takahashi et al. \(2012\)](#), as well as the neutrino corrections of [Bird et al. \(2012\)](#).

### 3.3.1 Galaxy observables

For the galaxy observables, we assume the flat sky and the Limber approximations (see Equation 1.42 and [Loverde and Afshordi 2008](#)). As such, we can simplify Equation 3.13 and obtain only one line of sight integral

$$C_\ell^{A,B} = \int \frac{c dz}{H(z) r^2(z)} W^A(k_\ell, z) W^B(k_\ell, z) P(k_\ell, z), \quad (3.14)$$

with  $k_\ell = (l + 1/2)/r(z)$ .

We assume that the projected galaxy density weight function is scale-independent. It is given by

$$W^{\text{GCP}_i}(z) = b_i(z) \frac{n_i(z) H(z)}{\bar{n}_i c}, \quad (3.15)$$

with  $n_i$  the number density of galaxies in the redshift bin  $i$ . This number density of galaxies is given by the convolution of the true redshift distribution of galaxies  $n(z)$  with the photometric redshift uncertainty (see [Kitching et al. 2009](#); [Euclid Collaboration et al. 2019](#), for more details).

We follow [Euclid Collaboration et al. \(2019\)](#) and assume that the galaxy bias  $b_i(z)$  is scale-independent and constant in each given redshift bin, with values  $b_i =$

$\sqrt{1+z_{c,i}}$  where  $z_{c,i}$  is the central redshift of the  $i$ th bin. The resulting 10 bias parameters  $b_i$  are part of the Fisher analysis as nuisance parameters.

The weak lensing convergence field is the distortion of the image of the background galaxies by the foreground matter. The weak lensing weight function includes contributions from both the cosmic shear signal ( $\gamma$ ) and the intrinsic alignment (IA) systematic effect. The tidal processes during the formation and evolution of galaxies may induce a preferred orientation of galaxy shapes. This intrinsic correlation of the orientation of galaxies is a contaminant of the shear two-point correlation function, and must be properly taken into account in a weak lensing survey.

The weak lensing weight function is expressed as

$$W^{\text{WL}_i}(z) = W_i^\gamma(z) - \frac{\mathcal{P}_{\text{IA}} \Omega_{\text{m},0}}{D(z)} W_i^{\text{IA}}(z). \quad (3.16)$$

The weight functions for the shear and intrinsic alignments are given by

$$W_i^\gamma(z) = \frac{3}{2} \frac{H_0^2}{c^2} \Omega_{\text{m},0} (1+z) r(z) \int_z^\infty dz' \frac{n_i(z')}{\bar{n}_i} \left[ 1 - \frac{r(z)}{r(z')} \right] \quad (3.17)$$

$$W_i^{\text{IA}}(z) = \frac{n_i(z)}{\bar{n}_i} \frac{H(z)}{c}. \quad (3.18)$$

The intrinsic alignment effect is modelled in  $\mathcal{P}_{\text{IA}}$ , which is described in [Euclid Collaboration et al. \(2019\)](#), and introduces three nuisance parameters in the Fisher analyses, namely  $\mathcal{A}_{\text{IA}}$ ,  $\eta_{\text{IA}}$  and  $\beta_{\text{IA}}$ , which correspond to the amplitude, the redshift dependence and the galaxy luminosity dependence of the IA power spectrum

### 3.3.2 CMB lensing

The primary image of the CMB is distorted by the gravitational lensing effects of the mass along the line of sight. We can recover the lensing potential signal by measuring the distortions in the primary image of the CMB. The CMB lensing efficiency peaks at a redshift of  $z \sim 2$ . Due to its late-time nature we can express its auto and cross angular power spectra using the Equation 4.10 in the Limber approximation and with the following weight function

$$W^\phi(z) = \frac{3}{2} H_0^2 \Omega_{\text{m},0} (1+z) r(z) \left[ 1 - \frac{r(z)}{r_*} \right], \quad (3.19)$$

where  $r_*$  is the comoving distance to the surface of last scattering (see e.g. [Lewis and Challinor 2006](#), for a review).

Note that the galaxy weak lensing and the CMB lensing have a similar weight function. The difference is that the source plane is at a known redshift for the CMB while we must integrate on the photometric redshift distribution of source galaxies for the galaxy weak lensing.

### 3.3.3 CMB temperature and polarisation

Computations of the power spectra of primordial observables such as the CMB temperature and polarisation cannot be performed with the Limber approximation. We should then rely on Equation 3.13 to compute their angular power spectra and their cross-correlation with the late time observables described above.

In the linear regime, we can decouple the redshift and scale dependence in the matter power spectrum and rewrite Equation 3.13 as

$$C_{\ell}^{A,B} = \frac{2}{\pi} \int k^2 dk \Delta_{\ell}^A(k) \Delta_{\ell}^B(k) P(k), \quad (3.20)$$

with  $\Delta_{\ell}(k)$  the kernels corresponding to each probe.

The CMB kernels take into account linear effects that create anisotropies in the CMB namely the Sachs-Wolfe effect, the Doppler effect, the integrated Sachs-Wolfe effect (iSW), the polarisation and the reionisation. We refer to [Ma and Bertschinger \(1995\)](#) and [Seljak and Zaldarriaga \(1996\)](#) for a detailed description of these kernels for the CMB temperature and polarisation observables.

The cross-correlation of temperature and polarisation with late time probes which are GCP, WL and  $\phi$  are obtained with Equation 3.20. The Boltzmann codes we use (CLASS and CAMB) implement specific recipes in order to take into account non linear effects at the kernel levels, which we do not detail here.

Note that in our analysis we do not take into account non-linear effects such as the Rees-Sciama effect, which is similar to the iSW but created by the non linear collapse of gravitational potential wells. We also do not include spectral distortions effects such as the tSZ effect, the Cosmic Infrared Background (CIB) and point sources. We assume that these contributions have been properly removed from the observed temperature and polarisation maps that are used in the analysis.

## 3.4 *Euclid* observables

### 3.4.1 Introducing the *Euclid* mission

The main goals of the *Euclid* mission are to understand the nature of the apparent expansion of the Universe and to test gravity on cosmological scales ([Laureijs et al. 2011](#); [Amendola et al. 2018](#)). To reach these goals, *Euclid* will measure the shape and redshift of galaxies over 15 000 deg<sup>2</sup> of the extragalactic sky up to  $z \sim 2.5$ . The survey will estimate the spectroscopic redshift of thirty millions of galaxies, and will get the image and photometric redshift of two billions of galaxies.

The current planned launch of the satellite is 2022, from the Kourou launchpad in a Soyouz rocket. It will reach the Lagrange L2 point, and will start mapping the sky shortly after that. The nominal mission duration is six years, with a possible one year extension. The data will be released in three batches: two, four and seven years after the launch.

The telescope has a mirror of 1.20 m of diameter and two instruments sharing a common field of view of 0.54°. The two instruments are the visible light imager (VIS) and the Near Infrared Spectrometer and Photometer (NISF).

The VIS instrument is an imager with a broad-band filter which has a wavelength range of 500 to 800 nm, and a pixel resolution of 0.1 arcseconds. The galaxy catalogue will be complete down to a magnitude of 24.5 AB in this band.

The NISP instruments has two wheels (a filter wheel and a grism wheel) in front of its detector that will allow to shift between a photometric and a spectroscopic mode. On the filter wheel are mounted three filters in Y, J, and H bands with a pixel resolution of 0.3 arcseconds. On the grism wheel are mounted four grism spectrographs, one ‘blue’ grism (920 – 1250 nm), and three ‘red’ grisms (1250 – 1850 nm) each with a different orientation. The galaxy catalogue will be complete to magnitude of 24 AB in the Y, J and H bands.

The grism will perform slitless spectroscopy, allowing for precise redshift determinations for about 1950 gal deg<sup>-2</sup>. The spectra of each galaxy will be directly diffracted on the detector. As such the image will be composed of the spectra of all the galaxies in the field of view, and the spectra of neighbouring galaxies may overlap. Taking the same image with different grism orientations will ensure the correct extraction of the spectra of each galaxy. The expected uncertainty on the spectroscopic redshift is of  $\sigma_z = 0.001(1 + z)$

The spectroscopic survey will target the  $H\alpha$  emission line of galaxies. This line is emitted at 656.3 nm in the galaxy rest frame, and corresponds to the emission of a photon when the electron of the hydrogen atom goes from the energy level  $n = 3$  to  $n = 2$ . This transition usually happens just after an ionised hydrogen atom recombines with an electron. This emission line is a good tracer of star forming galaxies, as the powerful young stars emitting in the UV will ionise the surrounding neutral hydrogen. We know that the star formation was higher at high redshift, and peaked at  $z \sim 2$  (Madau and Dickinson 2014). As such we expect to detect more  $H\alpha$  emitting galaxies when we observe galaxies in the redshift range  $0.9 < z < 1.8$ , which is the redshift window of the grisms (Pozzetti et al. 2016).

The photometric survey will allow to measure the redshift with a precision of  $\sigma_z = 0.05(1 + z)$ , by fitting a Spectral Energy Distribution (SED) template on the fluxes of the four bands of the survey. Additional data from ground based telescopes in other bands will help in reaching a high precision in the redshift estimate. When fitting an SED template there is a risk of catastrophic redshift estimation when the fit is completely wrong. In Euclid Collaboration et al. (2019), this outlier fraction is estimated to be of 10%.

Three fields of the sky representing a total of 40 deg<sup>2</sup> will be observed several times and will allow for a deeper survey. This deep survey will reach an AB magnitude of 26.5 in the VIS band and 26 in the near infrared bands.

### 3.4.2 Implementation of the *Euclid* observables

The WL and GCp tomographic analysis is performed in 10 equi-populated redshift bins, i.e. there is the same number of galaxy in each bin, between  $z = 0$  and  $z = 2.5$ . The underlying galaxy distribution is modelled by Laureijs et al. (2011)

$$n(z) \propto \left(\frac{z}{z_0}\right)^2 \exp\left[-\left(\frac{z}{z_0}\right)^{3/2}\right]; \quad (3.21)$$

Table 3.1: Specifications for the *Euclid* photometric survey.

	Parameter	<i>Euclid</i>
Survey area in the sky	$A_{\text{survey}}$	15,000 deg <sup>2</sup>
Sky fraction	$f_{\text{sky}}$	$\sim 0.36$
Galaxy number density	$\bar{n}_{\text{g}}$	30 arcmin <sup>-2</sup>
Total intrinsic ellipticity dispersion	$\sigma_{\epsilon}$	0.30
Minimum redshift	$z_{\text{min}}$	0.001
Maximum redshift	$z_{\text{max}}$	0.9 (pessimistic), 2.5 (optimistic)
Number of redshift bins	$N_z$	5 (pessimistic), 10 (optimistic)
Minimum multipole	$\ell_{\text{min}}$	10
Maximum multipole	$\ell_{\text{max}}$	1500 (pessimistic), 5000 (optimistic)
Number of bins in multipole space	$N_{\ell}$	100

with  $z_m = 0.9$  and a mean surface density of galaxies of  $\bar{n}_{\text{g}} = 30 \text{ arcmin}^{-2}$ .

The shot-noise terms for the angular power spectra are modelled by

$$N_{\ell}^{\text{GCp}_i, \text{GCp}_j} = \delta_{i,j}^{\text{K}} \frac{1}{\bar{n}_i}, \quad (3.22)$$

$$N_{\ell}^{\text{WL}_i, \text{WL}_j} = \delta_{i,j}^{\text{K}} \frac{\sigma_{\epsilon}}{\bar{n}_i}, \quad (3.23)$$

where  $\sigma_{\epsilon}$  is the total intrinsic ellipticity dispersion and  $\delta^{\text{K}}$  is the Kronecker delta. We assume that there is no noise for the cross-correlation between the weak lensing and the galaxy clustering.

In order to use the same hypotheses and recipes than in the analysis of [Euclid Collaboration et al. \(2019\)](#), our analysis considers two different settings for *Euclid* observables. These two settings are defined in order to test the amount of information we will extract from the *Euclid* probes, assuming that we will be able to perfectly reproduce the clustering of matter on the smallest scales, or that we will still be limited and will only extract information on large scales. We summarize below these two settings

**Pessimistic setting:** The spectroscopic galaxy clustering probe is performed out to  $k_{\text{max}} = 0.25 h \text{ Mpc}^{-1}$ . The maximum multipole of the angular power spectra is of  $\ell_{\text{max}} = 1500$  for WL, and of  $\ell_{\text{max}} = 750$  for GCp and the cross correlation between the two. We also cut the GCp tomography to include only the redshift bins  $z < z_{\text{max}} = 0.9$ , in order to avoid the overlap with the GCs probe. As such we only have 5 redshift bins for the GCp ranging from  $z = 0$  to  $z = 0.9$ .

**Otimistic setting:** The spectroscopic galaxy clustering probe is performed out to  $k_{\text{max}} = 0.30 h \text{ Mpc}^{-1}$ . The maximum multipole of the angular power spectra is of  $\ell_{\text{max}} = 5000$  for WL, and of  $\ell_{\text{max}} = 3000$  for GCp and the cross correlation between the two. The tomography of GCp is performed in the ten redshift bins.

The Table 3.1 summarizes the specifications of the *Euclid* WL and GCp observables.

The 3D spectroscopic galaxy clustering GCs is considered as an independent probe (thus uncorrelated with all other probes considered) throughout the whole analysis. The contribution of GCs to our final Fisher matrices is therefore accounted for via the simple addition of the official corresponding Fisher matrix published by [Euclid Collaboration et al. \(2019\)](#). A pessimistic and optimistic version of GCs is also considered.

## 3.5 CMB experiments and noise models

Similarly to Euclid observables, I have constructed the CMB noise models for the forecasts. We considered a variety of scenarios for the characteristics of the CMB observables: a *Planck*-like full extra-galactic sky survey and two types of ground-based observatories based on future experiments, namely the Simons Observatory (SO) and CMB Stage 4 (CMB-S4). In the latter two cases, the fraction of the sky expected to be covered will be of order 40% and 60% respectively. In order to avoid neglecting the precious amount of information contained in the CMB at large scales (and their exquisite measurements by *Planck*), the two ground-based scenarios will also include information on large scale based on signal observed by *Planck* (via a split in the multipole range in the Fisher analysis, cf. Table 3.2).

### 3.5.1 *Planck*

The *Planck* satellite ([Tauber et al. 2010](#); [Planck Collaboration et al. 2018c](#)) was launched in 2009 and scanned the full sky until 2013 in nine frequencies from 30 to 857GHz. The satellite hosted two instruments, the HFI operating in six frequency bands between 100 GHz and 857 GHz, and the LFI instrument operating in three band between 30 GHz and 77 GHz. The CMB maps were produced by combining these frequencies to remove the contributions from the Galaxy and other foreground emissions.

In our analysis, the noise of the CMB power spectra is modelled as an isotropic noise deconvolved by the instrument beam ([Knox 1995](#))

$$N_{\ell}^{\text{A,A}} = (\Delta\text{A})^2 b_{\ell}^{-2}, \quad (3.24)$$

$$b_{\ell} = \exp\left(\frac{-\ell(\ell+1)\theta_{\text{FWHM}}^2}{16 \ln 2}\right), \quad (3.25)$$

where  $\theta_{\text{FWHM}}$  is the full-width-at-half-maximum (FWHM) of the beam given in radian and  $\Delta\text{T}$  and  $\Delta\text{E}$  are the detector noise levels, for temperature and polarisation, given in  $\mu\text{K arcmin}$ . The total noise for multiple frequency channels is given by their inverse noise weighted sum. For *Planck*, this corresponds to a noise of  $\Delta\text{T} = 27 \mu\text{K arcmin}$  in temperature and  $\Delta\text{E} = 42 \mu\text{K arcmin}$  in polarisation. We use an effective beam size of  $\theta_{\text{FWHM}} = 7 \text{ arcmin}$ . We use a sky fraction  $f_{\text{sky}} = 0.7$  and a maximum multipole  $\ell_{\text{max}} = 1500$  for TT, TE and EE angular power spectra. In order to mimic the  $\Lambda\text{CDM}$  constraints given in [Planck Collaboration et al. \(2019\)](#), we multiply  $N_{\ell}^{\text{EE}}$  by a factor eight for  $\ell < 30$ .

The CMB lensing field has been estimated with a minimum variance quadratic estimator, combining temperature and polarisation data (Okamoto and Hu 2003). The *Planck* survey provided the most precise map of the integral of the density of matter on the full extra galactic sky, covering  $\sim 70\%$  of the sky, which allowed to get an estimate of the lensing-potential power spectrum over lensing multipoles  $8 \leq \ell \leq 400$ . The noise of the CMB lensing field reconstructed from *Planck* is taken from Planck Collaboration et al. (2018b).

### 3.5.2 Simons Observatory

The Simons Observatory (SO, The Simons Observatory Collaboration et al. 2019) consists of four different telescopes placed in the Atacama Desert in Chile, with the goal of providing an exquisite mapping of the CMB intensity and polarisation anisotropies from a few degrees down to arcminute scales. It is expected to start collecting data in 2022. Three of these telescopes have 0.5 m of aperture, and with an angular resolution close to half a degree, will map 10 % of the sky targeting the moderate to large angular scales. Their primary goal is to measure large-scale polarisation from the background of primordial gravitational waves.

Alongside these small telescopes, one 6 m diameter telescope will produce data appropriate for combination and cross-correlation with *Euclid*. It will observe at 27, 39, 93, 145, 225, and 280 GHz, with an angular resolution close to the arcminute and will reach a sensitivity level of  $6 \mu\text{K arcmin}$  on 40 % of the sky.

We use in our Fisher analysis the noise curves provided by the SO Collaboration<sup>1</sup>. In practice, we took the noise curves obtained with the internal linear combination (ILC) component separation method, assuming the baseline analysis for a sky fraction of 0.4. We differ here from the formula used for *Planck*, as this noise is modelled using the component separation method for all channels. For our forecasts with SO, we use data from  $40 \leq \ell \leq 3000$  for TT, TE and  $\phi\phi$ , and  $40 \leq \ell \leq 5000$  for EE. As mentioned at the beginning of the section, this scenario also considers the addition of large-scale data from the *Planck* survey; in practice, we add this information via the first multipoles of all CMB spectra considered, up to  $\ell = 40$ , with the same specifications as described earlier for the *Planck*-like survey.

### 3.5.3 CMB-Stage 4

The CMB-Stage 4 (CMB-S4, Abazajian et al. 2019) experiment will be the successor of the Simons Observatory, and will combine resources with the successor of the South Pole Telescope and the BICEP/Keck collaborations. It should start taking data in 2027. Its main scope is to measure the imprint of primordial gravitational waves on the CMB polarisation anisotropy, but it will also perform a wide survey with a high resolution which will allow to probe the secondary anisotropies with unprecedented accuracy. Its deep and wide survey will cover  $\sim 70\%$  of the sky and will be conducted over seven years using two 6 m telescopes located in Chile, each equipped with 121,760 detectors distributed over eight frequency bands from 30 GHz to 270 GHz. After removal of the galactic contamination, the sky fraction is

<sup>1</sup>We use the version 3.1.0 available at [https://github.com/simonsobs/so\\_noise\\_models](https://github.com/simonsobs/so_noise_models)

estimated to be between 40% and 60%. We choose to be conservative and to assume a 40% sky fraction. We model the noise power spectra as in Equation 3.24. These observations will provide CMB temperature and polarisation maps with a resolution of  $\theta_{\text{FWHM}} = 1$  arcmin and with noise levels of  $\Delta T = 1 \mu\text{K arcmin}$  and  $\Delta E = \sqrt{2} \mu\text{K arcmin}$ . CMB-S4 expects to use data from  $40 \leq \ell \leq 3000$  in temperature, and  $40 \leq \ell \leq 5000$  in polarisation.

The lensing noise curve is taken as the minimum variance (N0) bias, which is computed using the code `quicklens`<sup>2</sup>. Similarly to the SO case, *Planck*-like large-scale information is also added in this scenario.

All specifications for our three considered CMB experiments are summarized in Table 3.2, while the noise curves for all CMB auto-spectra are shown in Fig. 3.2.

Table 3.2: Specifications for CMB experiments

	Parameter	<i>Planck</i>	Simons Observatory + <i>Planck</i> low- $\ell$	CMB+Stage 4 + <i>Planck</i> low- $\ell$
Sky fraction	$f_{\text{sky}}$	0.7	0.4	0.4
Beam FWHM	$\theta_{\text{FWHM}}$	7 arcmin	2 arcmin	1 arcmin
Temperature noise	$\Delta T$	23 $\mu\text{K.arcmin}$	3 $\mu\text{K.arcmin}$	1 $\mu\text{K.arcmin}$
polarisation noise	$\Delta E$	42 $\mu\text{K.arcmin}$	$3\sqrt{2} \mu\text{K.arcmin}$	$\sqrt{2} \mu\text{K.arcmin}$
TT multipole range	$[\ell_{TT,\text{min}}, \ell_{TT,\text{max}}]$	[2, 1500]	[2, 39] + [40, 3000]	[2, 39] + [40, 3000]
TE multipole range	$[\ell_{TE,\text{min}}, \ell_{TE,\text{max}}]$	[2, 1500]	[2, 39] + [40, 3000]	[2, 39] + [40, 3000]
EE multipole range	$[\ell_{EE,\text{min}}, \ell_{EE,\text{max}}]$	[2, 1500]	[2, 39] + [40, 5000]	[2, 39] + [40, 5000]
$\phi\phi$ multipole range	$[\ell_{\phi\phi,\text{min}}, \ell_{\phi\phi,\text{max}}]$	[8, 400]	[2, 39] + [40, 3000]	[2, 39] + [40, 3000]
$T\phi$ multipole range	$[\ell_{T\phi,\text{min}}, \ell_{T\phi,\text{max}}]$	[8, 400]	[2, 39] + [40, 3000]	[2, 39] + [40, 3000]

## 3.6 Forecasting cosmological constraints

We describe in this Section the implementation of the Fisher analysis to forecast constraints for the joint analysis of *Euclid* and CMB observables. Our probes are the two-point angular power spectra, containing the auto and cross correlations between the different observables. We tested different cosmological models and different scenarios of probe combination (including or not the cross-correlation information between *Euclid* and CMB observables).

### 3.6.1 Extensions of $\Lambda\text{CDM}$

In the following, we assume that the neutrinos have a mass of  $\sum m_\nu = 0.06 \text{ eV}$ . Our baseline  $\Lambda\text{CDM}$  model is described by six free parameters: the reduced Hubble parameter ( $h = H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ), the density of matter ( $\Omega_m$ ), and of baryons ( $\Omega_b$ ), the spectral index of the primordial power spectrum ( $n_s$ ), the amplitude of the matter fluctuation smoothed by a top-hat window of 8 Mpc ( $\sigma_8$ ), and the reionization optical depth ( $\tau$ ).

Furthermore, we introduce several extensions to this baseline  $\Lambda\text{CDM}$  model.

<sup>2</sup><https://github.com/dhanson/quicklens>

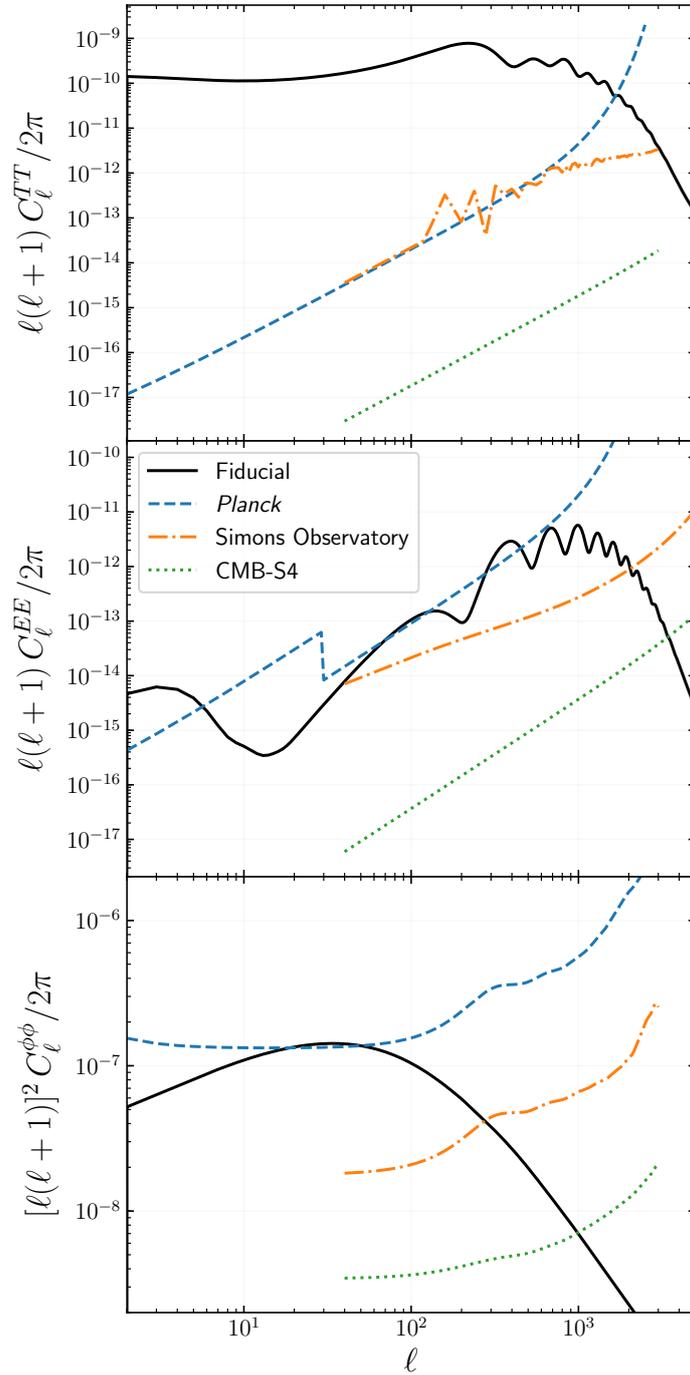


Figure 3.2: Noise curves of the CMB experiments under consideration. Black plain lines show the fiducial power spectra, dashed blue lines show the *Planck* noises, dot-dashed lines show the Simons Observatory noises and dotted green lines are the CMB-S4. Top panel is the Temperature power spectrum, central panel is the CMB E mode polarisation, and bottom panel is for the CMB lensing.

- The first one assumes that the dark energy equation of state evolves with time. Instead of having a constant dark energy equation of state parameter  $\omega_{\text{DE}} = -1$ , we let  $\omega_{\text{DE}}$  evolve with time. The CPL parametrisation (Chevallier and Polarski 2001; Linder 2003) assumes  $\omega_{\text{DE}}(a) = \omega_0 + \omega_a(1 - a)$ , with  $\omega_0$  its present day value and  $\omega_a$  a constant to describe its time evolution. In the following, we call this model  $\omega\text{CDM}$ , and  $w_0, w_a$  are taken as free parameters.
- The second extension allows for non-flat cosmologies. We consider  $\Omega_{\text{K}}$  as a free parameter of the model. Since, by definition,  $\Omega_{\text{K}} + \Omega_{\Lambda} + \Omega_{\text{m}} = 1$ , when we explore non-flat models  $\Omega_{\Lambda}$  is also varied in conjunction with  $\Omega_{\text{K}}$  (so as to keep all other parameters constant).
- The last extension modifies the  $\gamma$  parameter of the growth factor defined in Equation 1.49, in order to test deviations from general relativity. In this model  $\gamma$  is considered as a free parameter.

We can test each extension separately or we can combine them. We thus have eight possible cosmological models:  $\Lambda\text{CDM}$ ,  $w\text{CDM}$ , both flat or curved, and with  $\gamma$  fixed or free. Table 3.3 gives the fiducial values of the parameters considered in the baseline  $\Lambda\text{CDM}$  model and its extensions.

Table 3.3: Fiducial parameter values for the free parameters of our fiducial cosmological model, both in the baseline  $\Lambda\text{CDM}$  case and in its extensions.

Baseline						Extensions			
$\Omega_{\text{b}}$	$\Omega_{\text{m}}$	$h$	$n_{\text{s}}$	$\sigma_8$	$\tau$	$\Omega_{\text{DE},0}$	$w_0$	$w_a$	$\gamma$
0.05	0.32	0.67	0.96	0.816	0.058	0.68	-1	0	6/11

### 3.6.2 Fisher matrix

We assume that there is no correlation between different multipoles. Therefore, we compute the Fisher matrix for each multipole, and the total Fisher matrix is the sum on all multipoles

$$F_{\alpha\beta} = \sum_{\ell} \frac{\partial \mathbf{D}(\ell)}{\partial \theta_{\alpha}} \mathbf{C}^{-1} \frac{\partial \mathbf{D}(\ell)}{\partial \theta_{\beta}}. \quad (3.26)$$

The data vector  $\mathbf{D}(\ell)$  contains the auto and cross angular power spectra  $C_{\ell}^{\text{A,B}}$  of the *Euclid* and CMB observables, for a given multipole  $\ell$ .

The covariance matrix  $\mathbf{C}$  between the  $C_{\ell}^{\text{A,B}}$  is assumed to be Gaussian

$$\begin{aligned} \mathbf{C}_{\ell} \left( C_{\ell}^{\text{A,B}}, C_{\ell}^{\text{C,D}} \right) = & \frac{1}{(2\ell + 1) \Delta\ell f_{\text{sky}}} \left[ \left( C_{\ell}^{\text{A,C}} + \delta_{\text{A,C}}^{\text{K}} N_{\ell}^{\text{A,A}} \right) \left( C_{\ell}^{\text{B,D}} + \delta_{\text{B,D}}^{\text{K}} N_{\ell}^{\text{B,B}} \right) \right. \\ & \left. + \left( C_{\ell}^{\text{A,D}} + \delta_{\text{A,D}}^{\text{K}} N_{\ell}^{\text{A,A}} \right) \left( C_{\ell}^{\text{B,C}} + \delta_{\text{B,C}}^{\text{K}} N_{\ell}^{\text{B,B}} \right) \right], \end{aligned} \quad (3.27)$$

where the noise power spectra  $N_{\ell}$  are described in Sections 3.4.2 and 3.5

### 3.6.3 Scenarios for the combined analysis

We tested different configurations for the combination of the 2D *Euclid* observables (WL and GCp) to the CMB observables (T, E and  $\phi$ ). In all cases, we do not take into account the covariances between the 3D galaxy clustering probe, which we assume independent of the other 2D probes, as in (Euclid Collaboration et al. 2019).

In the first configuration, we assume that the CMB and *Euclid* observables are independent. We thus did not include the cross power spectra between the *Euclid* and CMB observables in the analysis, and we neglected the covariance between the two types of probes in our covariance matrix (see Figure 3.3 a). In practice, this is equivalent to summing the Fisher matrix of *Euclid* with that of the CMB to obtain combined constraints assuming two completely independent datasets.

In the second configuration, we do not assume that *Euclid* and CMB probes are independent anymore, and we properly take into account the covariance between them in the covariance matrix (see Figure 3.3 b). However, we do not include cross-power spectra in the data vector.

In the last configuration, we take into account in our data vector the probes that are the cross-correlations between the CMB and *Euclid* observables. We thus include the cross correlations between the temperature, polarisation and lensing of the CMB with the galaxy density and the galaxy weak lensing of *Euclid* (see Figure 3.3 c).

We show in Figure 3.3 the covariance matrices for the data vector  $\mathbf{D}(\ell = 83)$  in the three configurations of probes combinations.

As we can see in Figure 3.3 (a), the first configuration assumes that there is no correlation between the *Euclid* and CMB probes, the covariance matrix is composed of two independent blocks (the top left and bottom right blocks) corresponding to the covariance between CMB probes and *Euclid* probes only. The second configuration shown in Figure 3.3 (b) includes the correlations between *Euclid* and CMB probes (upper right and bottom left rectangles are now filled). Finally, the Figure 3.3 (c) shows the third configuration with the inclusion of the cross-correlations between *Euclid* and CMB observables in the data vector.

### 3.6.4 Limits of the numerical resolution

To test if the numerical analysis is stable and robust to the numerical uncertainties that can affect our analysis, I have performed stability analysis. We can see in Figure 3.3 that the values of the covariance matrices can range over 30 orders of magnitude. The computations in the Fisher matrix in Equation 3.26 relies on the inversion of these covariance matrices. As such, the numerical operation of the inversion could become very noisy if the resolution of the numerical inversion is not precise enough.

The condition number is a measure of the numerical accuracy needed for the inversion of the matrix. It gives an estimate of how much a small difference on the values in the covariance matrix  $\mathbf{C}$  will affect the values in the inverse matrix  $\mathbf{C}^{-1}$ .

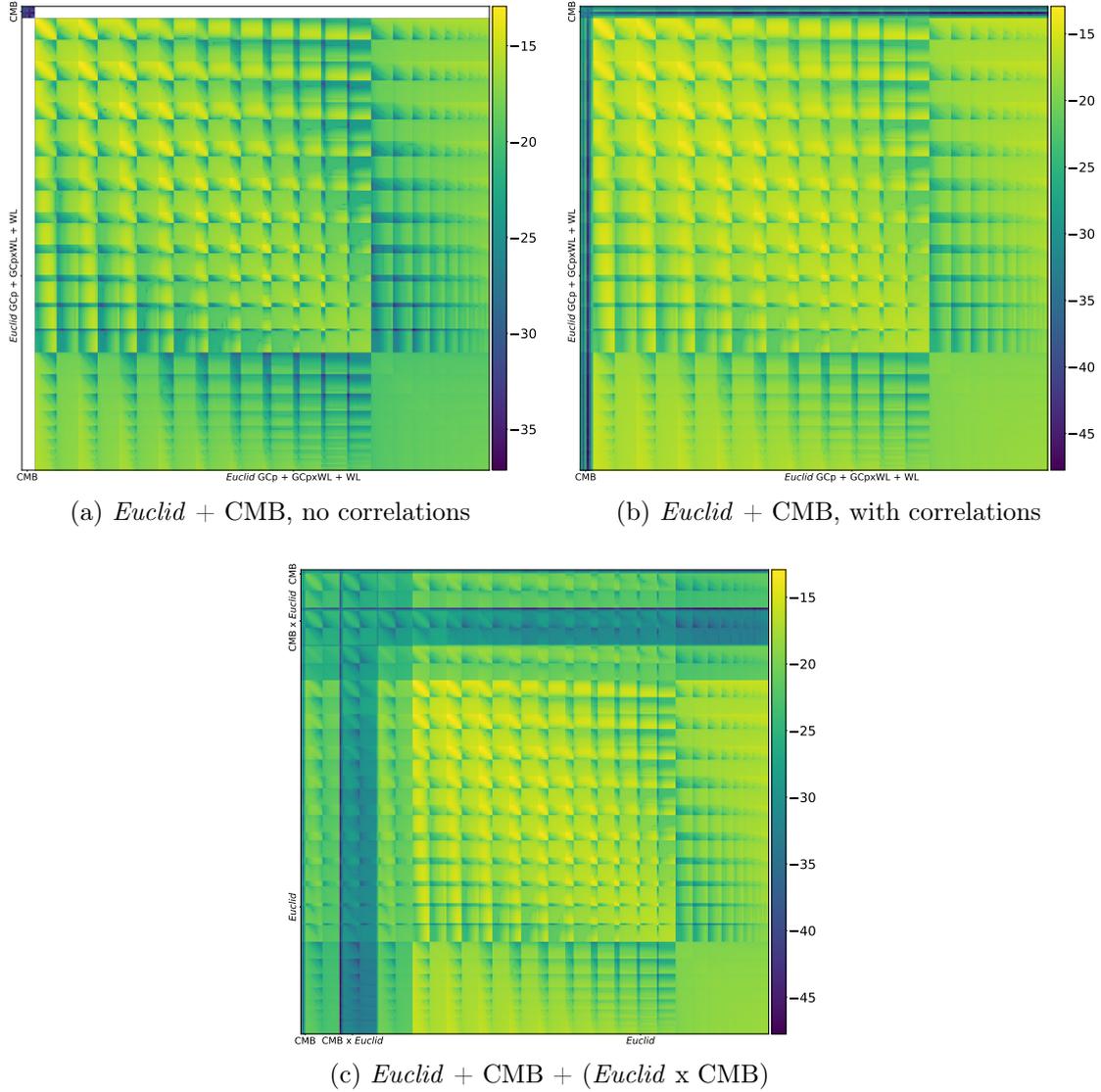


Figure 3.3: Logarithm of the absolute value of the covariance matrix for  $\ell = 83$  between the auto and cross angular power spectra of the observables T, E,  $\phi$ ,  $\text{GCp}_{i=1..10}$  and  $\text{WL}_{i=1..10}$ . We assume three different configurations of probe combination. *Top left*: no correlations between CMB and *Euclid* probes. *Top right*: including correlations between CMB and *Euclid* probes. *Bottom*: including cross-power spectra between *Euclid* and CMB observables in the data vector. The labels on the axis of each panel show roughly the location of CMB and *Euclid* probes, and in panel (c) the cross between the two. As we can see the values in the covariance matrix can span more than 30 orders of magnitudes. The smallest matrices (panels a and b) have a shape of 216x216 and the largest (panel c) have a shape of 276x276.

In our case, the condition number is defined as

$$\kappa(\mathbf{C}) = \frac{|\lambda_{\max}(\mathbf{C})|}{|\lambda_{\min}(\mathbf{C})|}, \quad (3.28)$$

where  $\lambda_{\max}$  and  $\lambda_{\min}$  are the largest and lowest (in terms of absolute value) eigenvalues of the matrix  $\mathbf{C}$ . If the condition number is very large, the inversion could be numerically noisy. On the contrary if it is close to one then it is easily invertible.

When solving the linear system  $\mathbf{A}x = \mathbf{B}$ , with  $x$  the unknown solution of the system, one can show that we have

$$\frac{\|\delta x\|}{\|x\|} \leq \kappa(\mathbf{A}) \frac{\|\delta b\|}{\|b\|}, \quad (3.29)$$

with  $\delta b$  the error on the matrix  $b$  and  $\delta x$  the error on the solution  $x$ .

In our case, we use the `numpy` package in `python`. The precision of the `float64` data type is of  $2.22 \times 10^{-16}$ . This precision is defined as the difference between 1.0 and the next smallest representable float larger than 1.0, so it is the numerical uncertainty  $\delta b$ .

The maximum condition number we have in our covariance matrices was of  $10^{36}$ . This means that our matrix inversion is very sensitive to noise, and the upper bound of the relative error on the inverted matrix is

$$\frac{\|\delta \mathbf{C}^{-1}\|}{\|\mathbf{C}^{-1}\|} \leq 10^{20}, \quad (3.30)$$

which is quite large.

To test the validity of the inversion, we reproduced our Fisher analysis but we introduced noise in the covariance matrix of the order of the precision of `numpy`. We found that the Fisher analysis is robust to this noise, with results changing by less than 1%. We see that our inversion is not limited by the very large condition number, and is robust to numerical noise.

Moreover, tests have been performed in our *Euclid* working group to use a `python` module which increases the resolution of the floats. We found out that the results stayed the same when using this increased precision.

## 3.7 Results

This section presents the results we obtained in our analysis. Note that these preliminary results are confidential and that they have not yet been presented to the *Euclid* publication board. The results shown here use results from the Fisher matrix code I developed during my PhD and from a second code developed in parallel in the CMBX working group. We have checked that our two Fisher codes give consistent results, with a variation of  $\pm 5\%$ .

In total, with the two *Euclid* settings (see Section 3.4.2) and the three CMB experiments we considered, we have six different surveys combinations. We also considered three different configurations of probe combinations (see Section 3.6.3). We considered eight different cosmological models:  $\Lambda$ CDM and  $w$ CDM, both with  $\gamma$  fixed or free, and both flat and non-flat (see Section 3.6.1). In total, we obtain 144 possible Fisher matrices, each with six to ten cosmological parameters and with eight to thirteen astrophysical systematic parameters.

We first highlight some of our results in Figure 3.4. We show here the ratio

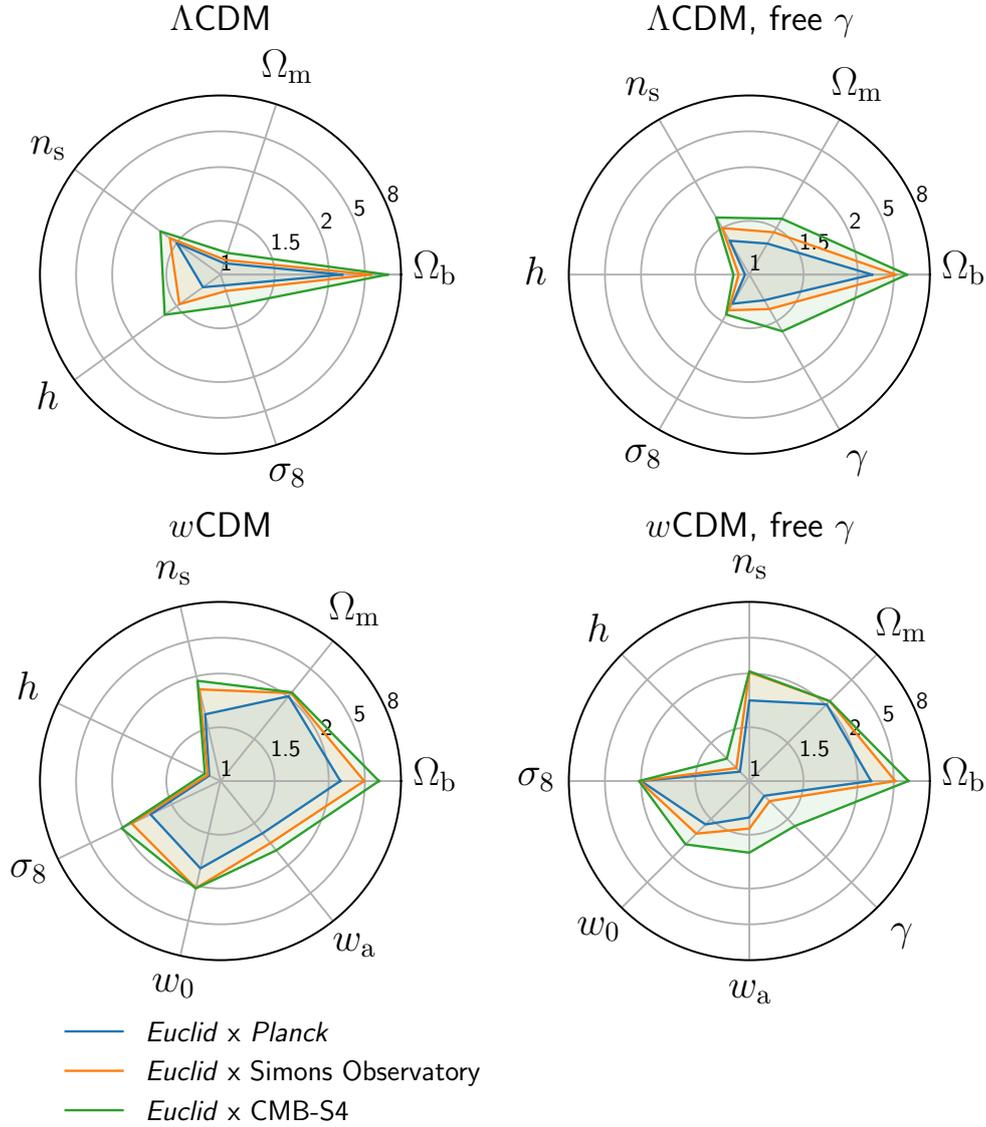


Figure 3.4: Ratio of the  $1\sigma$  marginalised uncertainties from *Euclid* only over the  $1\sigma$  marginalised uncertainties from *Euclid* x CMB. The *Euclid* x CMB constraints include the cross-correlated probes between the two datasets, and we include the GCs constraints in both cases. The radial axis is increased between 1 and 2 to better show small variations. We consider an optimistic *Euclid* experiment and a flat cosmology. Blue contours are when combining with *Planck*, orange contours are for Simons Observatory and green contours are for CMB-S4. Top panels assume a  $\Lambda$ CDM cosmology, bottom panels assume a  $w$ CDM cosmology, and right panels consider  $\gamma$  as a free parameter. We clearly see that  $\Omega_b$  constraints are greatly improved in all cases, and that it is on the  $w$ CDM models that the combination with CMB observables gives more improvements.

of the  $1\sigma$  marginalised uncertainties on the cosmological parameters from *Euclid* probes only over the  $1\sigma$  marginalised uncertainties from *Euclid* x CMB, including the cross-correlated probes between the two datasets (configuration 3 of Section 3.6.3). This shows the factor of improvement gained on the marginalised constraints when performing the full joint analysis between the two datasets. In this Figure, we show different results for flat cosmologies, and assuming an optimistic *Euclid* setting.

In all models shown here,  $\Omega_b$  is the parameter that is best improved by the combination with CMB datasets, with a factor two improvement for *Planck* and a factor eight for CMB-S4 in the  $\Lambda$ CDM cosmology. On the contrary,  $h$  is the parameter that is the least improved, with only a  $\sim 10\%$  improvements in the extended models. In this Figure, we clearly see that it is for extended cosmological models that the combination with CMB brings the most information. Except for  $h$ , all parameters have constraints improved by a factor of  $\gtrsim 2$  in the wCDM model (bottom left panel), and by a factor of  $\gtrsim 1.5$  in the  $\Lambda$ CDM model when  $\gamma$  is considered as a free parameter (upper right panel). The constraints on  $w_0, w_a$  in the wCDM model are improved by at least a factor 1.5, and up to two, for the three CMB experiments considered here. This means an increase in the figure of merit of  $w_0, w_a$  by a factor two to four.

Of course, the best constraints are achieved by the combination with the CMB-S4 experiment. However, it seems that the Simons Observatory will already greatly improve the constraints in most of the cases. Even *Planck* alone performs well on parameters of the wCDM model, by improving constraints on all parameters except  $h$  by more than 50%.

Figure 3.5 summarizes most of our results. This Figure displays the ratio of the  $1\sigma$  marginalised uncertainties for the *Euclid* observables alone over the constraints for *Euclid* + CMB combined analysis. The *Euclid* + CMB case is the second configuration described in Section 3.6.3 i.e. we do not take into account the cross-correlated probes in our data vector but we properly take into account the covariances between the two datasets. The colours correspond to the percentages of improvement, while the numbers in the matrix correspond to the factors of improvement.

We see that there is no improvements on the intrinsic alignment parameters in all the models considered here. It appears also that the addition of CMB constraints is more helpful for the pessimistic *Euclid* scenario than for the optimistic one. This is expected as the pessimistic scenario gives much less constraints on parameters than the optimistic one.

The parameters that benefit the most from the addition of CMB constraints are  $\Omega_b$  and  $\Omega_\Lambda$  (for relevant models), which are improved by a factor three to 13.7 depending on the model and setting considered. On the contrary, it seems like  $h$  is the parameter which is the least improved by the addition of CMB constraints. It is also clear, again as expected, that it is with CMB-S4 that we will get the highest improvements.

We show in Figure 3.6 the improvements when including the cross-correlated probes between *Euclid* and CMB in the data vector as opposed to not including them. This time both the colours and the numbers in the matrix are the percentage of improvements. This represents the improvement of including cross-correlated

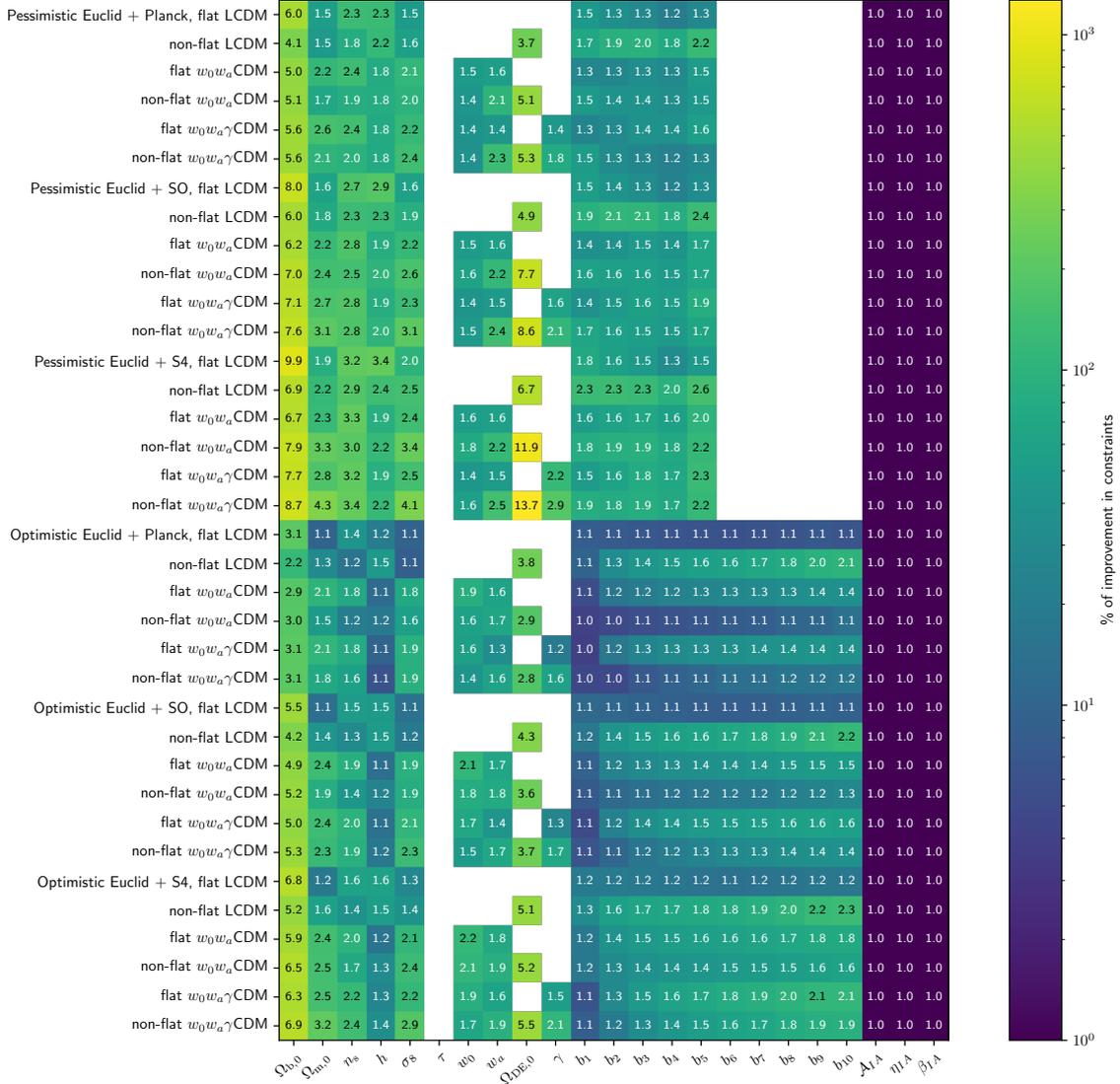


Figure 3.5: Figure from *Euclid* Collaboration (in prep.). Summary of our results, showing the ratio of the  $1\sigma$  marginalised uncertainties on the parameters of the models considered of the *Euclid* only case over the *Euclid* combined with CMB case, neglecting correlations between the two datasets.

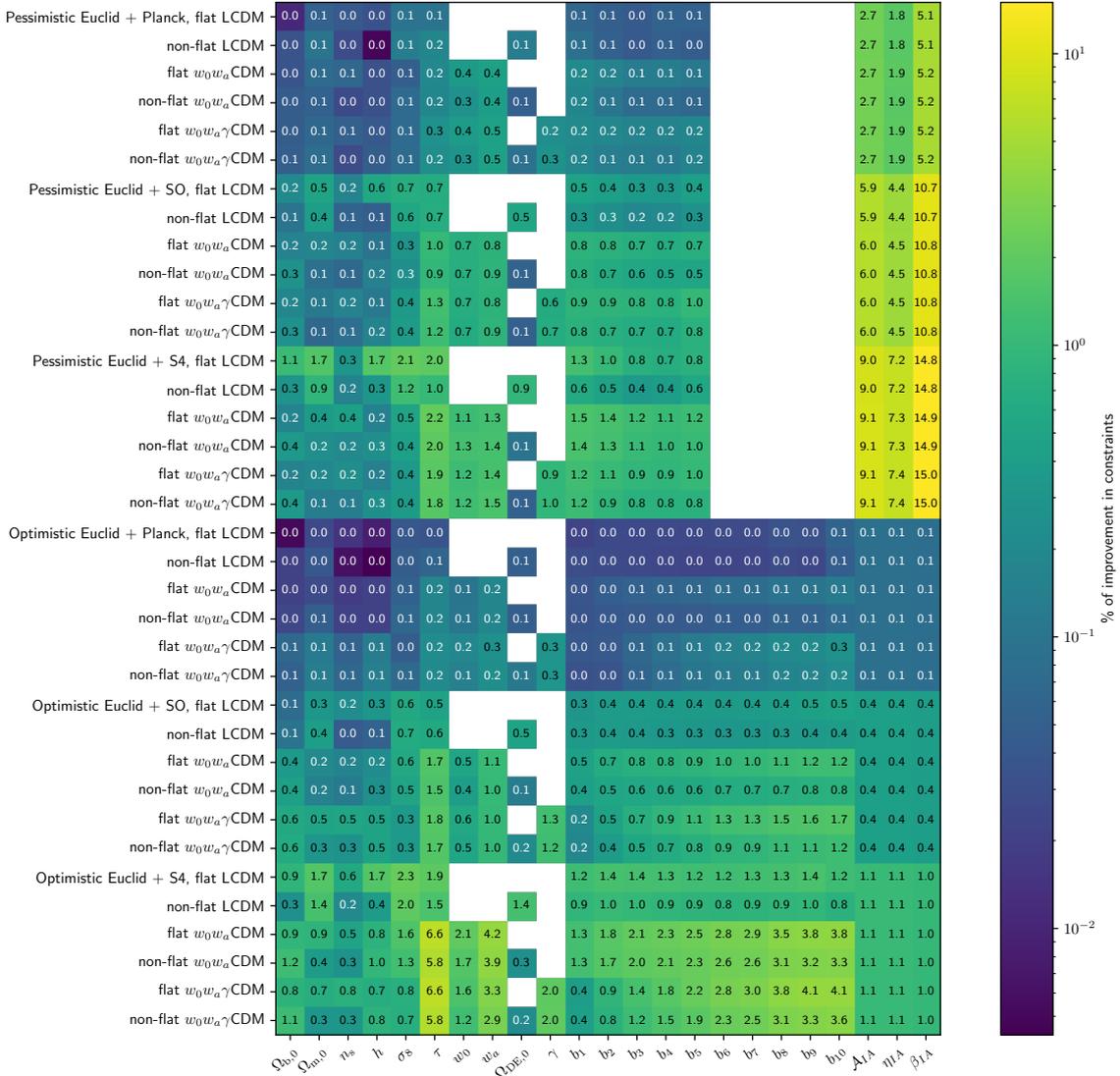


Figure 3.6: Improvements on marginalised  $1\sigma$  constraints after including cross-correlated probes *Euclid* x CMB as opposed to not including them in the *Euclid* + CMB combination. The colours and the numbers in the matrix show the percentage of improvements. The *Euclid* + CMB combination (our reference here) includes the correlations between observables in the covariance matrix.

probes on top of the factor of improvement when combining *Euclid* and CMB probes.

We see that these cross-correlated observables are improving the constraints on the intrinsic alignments parameters in case of a pessimistic *Euclid* scenario, with improvements on  $\mathcal{A}_{\text{IA}}$ ,  $\eta_{\text{IA}}$  and  $\beta_{\text{IA}}$  of 2.7%, 1.8%, 5.1% for *Planck* and of 9.1%, 7.4%, 15% for CMB-S4, and for all cosmological models.

Except for IA, we see that the cross-correlation with *Planck* observables do not bring great improvements, with a maximum of 0.5% improvement for a pessimistic *Euclid* setting and a maximum of 0.3% improvement for an optimistic *Euclid* setting.

We see that for SO and CMB-S4, the use of cross-correlated observables is more interesting in the optimistic *Euclid* setting. Indeed in this setting, we use the highest redshift bins of galaxy clustering photometric, between  $z = 0.9$  and  $z = 2.5$ , so the cross-correlation of galaxy density and CMB lensing is probing the peak of the CMB lensing efficiency ( $z \sim 2$ , see Equation 3.19). It appears that the galaxy bias parameters are improved by up to 4% when cross correlating CMB-S4 and *Euclid* observables. The optical depth of reionization parameter also benefits from the cross-correlation of *Euclid* and CMB, with an improvement of up to 6.6% with CMB-S4.

Finally, we compare in Figure 3.7 the constraints when taking into account or not the covariances between *Euclid* and CMB probes. This case is for the non flat  $\Lambda$ CDM model where  $\gamma$  is free, and combining *Euclid* optimistic with CMB-S4. It appears that the difference on the constraints when taking or not into account the covariances between CMB and *Euclid* probes is very marginal, with less than 1% of difference. It is probably due to the very low correlations between *Euclid* and CMB probes. However, to perform a precise analysis, one has to properly take into account the covariances between the probes, even if it appears here that the difference might be marginal.

## 3.8 Conclusion

I have participated to the production of the forecasts for the joint analysis of *Euclid* and CMB observables, as part of my collaboration inside the CMBX science working group. Specifically, I was involved in the modelling of CMB noise power spectra, the implementation and production of the covariance matrices between our probes, the implementation of the Fisher matrices analyses with numerical stability tests, and in the writing of the paper.

In this collaboration, we have produced forecasts for the constraints achieved by the combination of the *Euclid* main probes, namely galaxy clustering spectroscopic, galaxy weak lensing and galaxy clustering photometric, with CMB observables, namely temperature, polarisation and lensing potential. We used the Fisher formalism to forecast *a priori* how well the integration of CMB observables will improve the constraints given by *Euclid* probes. We have tested two *Euclid* settings and three CMB experiments over several extended cosmological models. We included astrophysical systematics, namely the galaxy bias and the weak lensing intrinsic alignments.

We found that a combined analysis of *Euclid* and CMB probes will greatly improve constraints on cosmological parameters, especially on extensions of the  $\Lambda$ CDM

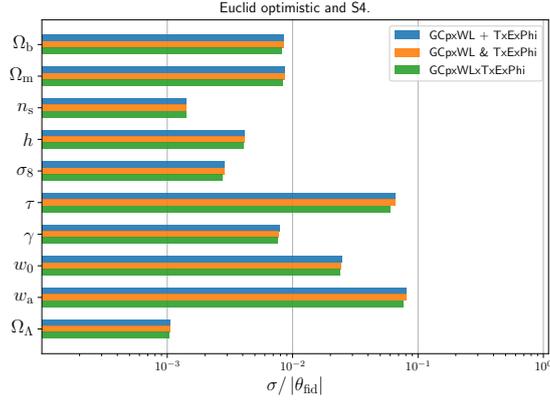


Figure 3.7: Ratio of the  $1\sigma$  marginalised uncertainty over the fiducial values of the parameters of the non flat  $w$ CDM model where  $\gamma$  is let free, for a *Euclid* optimistic and a CMB-S4 scenario. The three colours correspond to the three configuration of probe combination described in Section 3.6.3. The blue bars correspond to not taking into account the covariance between *Euclid* and CMB probes, the orange bars correspond to properly taking their covariances, and the green bars correspond to including the cross-correlations between *Euclid* and CMB observables in the data vector.

model. For instance, we can expect a factor two improvement in the  $w_0, w_a$  figure of merit for the combination of *Euclid* and *Planck*, and a factor four improvement for the combination with CMB-S4.

The parameter which benefits most from the combination with CMB data is  $\Omega_b$ . This may be due to the fact that the galaxy survey probes are sensitive to the total matter (dark matter and baryons) and not specifically to baryons, while the shape of the CMB power spectrum is greatly sensitive to density of baryons.

We have seen that the inclusion of cross-correlated probes seems to give marginal improvements. This small improvement ( $< 10\%$  compared to not including the cross-correlations) principally benefits to astrophysical systematic parameters such as the galaxy bias and the intrinsic alignments but also to the reionisation optical depth. The fact that the reionisation optical depth parameter is improved when using cross-correlated observables might come from the degeneracy between  $\tau$  and the amplitude of perturbations  $A_s$ , which is broken when adding the tight constraints on  $\sigma_8$  from large scale surveys.

In order to explore how to fully benefit from the power of the cross-correlated probes, we probably have to include more systematic effects in our settings, such as instrumental systematics or other sources of contaminations of the signal (residual galaxy emissions in the CMB for instance). Indeed, these types of systematics specific to each observable should disappear in the cross-correlations.

Our analysis is greatly simplified compared to a real case scenario. Indeed we did not include the effects of masking, we assumed a totally gaussian covariance between the probes, without correlations between different multipoles, and we neglected the super sample covariance. We also neglected the covariance between the 3D and 2D galaxy power spectra. In future studies, especially when the real data will arrive,

we will have to include all these effects in the likelihood analysis.

Our results are however very promising as we showed that the addition of CMB probes can help in improving the constraints on most parameters, especially for extended cosmological models. As the *Euclid* survey is specifically dedicated to explore extensions of  $\Lambda$ CDM, we argue that the combination of *Euclid* observables with CMB datasets will greatly help reaching this goal. A combined analysis of both datasets will help us in making the distinction between different cosmological scenarios.

# Chapter 4

## An alternative probe for galaxy surveys

The previous chapter presented the forecasts on the constraints obtained by the joint analysis of the *Euclid* survey with CMB surveys. In this context we used the main *Euclid* probes which are the weak lensing and the galaxy clustering (2D and 3D).

In order to maximise the cosmological power of next generation galaxy surveys, new probes of the distribution of matter are being developed. In this chapter we forecast the cosmological power of the angular redshift fluctuation (ARF). We describe the modelling of this probe and we forecast the constraints it will achieve on cosmological parameters. The work presented here has been submitted for publication (Legrand et al. 2020).

### 4.1 Introduction

In spectroscopic galaxy surveys, it is customary to convert redshift estimates into radial distances under the assumption of a given fiducial cosmological model. Angular and redshift coordinates are thus converted into the three-dimensional space, where standard 3D clustering analysis techniques are applied.

Here we choose to follow a different strategy. We focus on a new cosmological observable, the angular redshift fluctuations (ARF) which has been introduced in Hernández-Monteagudo et al. (2019). We explore its sensitivity to cosmology, either when considered independently, or when combining it with other *angular* observables, such as the standard angular clustering, or the lensing of the CMB. Being a 2D observable, the ARF field can easily be cross-correlated with other 2D observables, such as the 2D galaxy density field and the CMB lensing fields. ARF present other interesting features, such as being correlated to the cosmic, radial, peculiar velocity fields (Hernández-Monteagudo et al. 2019; Chaves-Montero et al. 2019), or being particularly insensitive to additive systematics that remain constant under the redshift shell subject to analysis.

We apply the Fisher formalism on the angular galaxy clustering, the ARF and the CMB lensing convergence observables, and explore their sensitivity to cosmology in two different observational set ups, mimicking those expected for the DESI and *Euclid* surveys. We consider the CMB lensing convergence field among our observables, since it constitutes an intrinsically different probe whose dependence on the

parameters defining the galaxy sample is different from that of angular galaxy clustering and ARF. Our scope is to assess if the ARF field can provide complementary information on the galaxy density field and on the CMB lensing field.

Throughout this chapter, we use the *Planck* 2018 cosmology as our fiducial cosmology. We take the values given in Table 2, in the column 6 (best-fit with BAO), of [Planck Collaboration et al. \(2018a\)](#). As introduced in Section 1.7.2, the redshift due to the Hubble expansion is denoted by  $z$ , while  $z_{\text{obs}}$  is the measured redshift (which includes redshift distortions induced by radial peculiar velocities). In this Chapter,  $r(z) = \int dz c/H(z)$  is the line of sight comoving distance, and  $dV_{\Omega} = dV/d\Omega = r^2 dr = r^2(z) c/H(z) dz$  is the comoving volume element per solid angle, with  $d\Omega$  a differential element in solid angle.

## 4.2 Surveys under consideration

Among the wealth of current and upcoming experiments, we choose two representative cases for spectroscopic large scale structures (LSS) surveys, namely the DESI and the *Euclid* experiments. We detail their specifications in Table 4.1. On the CMB side we consider the same experiments which were introduced in Section 3.5, namely a *Planck*-like experiment, the Simons Observatory and the CMB-Stage 4. We introduce below the specifications for the two galaxy surveys.

### 4.2.1 The DESI experiment

DESI is a ground-based survey that will cover 14,000 deg<sup>2</sup> on the sky and will measure the redshift of about 30 million galaxies using optical fibers spectroscopy ([DESI Collaboration et al. 2016](#)). It will target four different classes of galaxies. In this chapter, we compute forecasts for the emission line galaxies (ELG) sample which is the largest sample of the survey. It ranges from  $z = 0.6$  up to  $z = 1.6$ . The expected galaxy distribution  $\bar{n}_g(z)$  (see Figure 4.1a) and the galaxy bias  $b(z)$  are calibrated based on the DEEP2 survey ([Newman et al. 2013](#)). The (linear) bias of the spatial distribution of this galaxy population with respect to dark matter is a redshift dependent quantity approximated by

$$b_g(z) = 0.84/D(z) , \quad (4.1)$$

with  $D(z)$  denoting the growth factor of linear matter density perturbations.

The DESI experiment successfully achieved the commissioning phase in the first half of 2020 and should start observing the sky in the second half of 2020 for a five year period. The DESI collaboration plans to release annual datasets.

### 4.2.2 The *Euclid* spectroscopic survey

We summarize here the key points that have been introduced already in Section 3.4. We are now focusing on the spectroscopic survey which will be performed with the NISP instrument with slitless spectroscopy. This will allow for precise redshift determinations for about 1950 gal deg<sup>-2</sup> in the redshift range  $0.9 < z < 1.8$ . We

Survey	<i>Euclid</i>	DESI
Survey area	15 000 deg <sup>2</sup>	14 000 deg <sup>2</sup>
Redshift estimation	Slitless spectroscopy	Optical fiber spectroscopy
Targets	$H_\alpha$ emission line	[OII] doublet
Redshift range	$0.9 < z < 1.8$	$0.6 < z < 1.6$
Galaxy density	1950 gal deg <sup>-2</sup>	1220 gal deg <sup>-2</sup>
Galaxy bias	$b_g(z) = 0.79 + 0.68 z$	$b_g(z) = 0.84/D(z)$
Reference	<a href="#">Euclid Collaboration et al. (2019)</a>	<a href="#">DESI Collaboration et al. (2016)</a>

Table 4.1: Specifications for the two galaxy surveys under consideration.

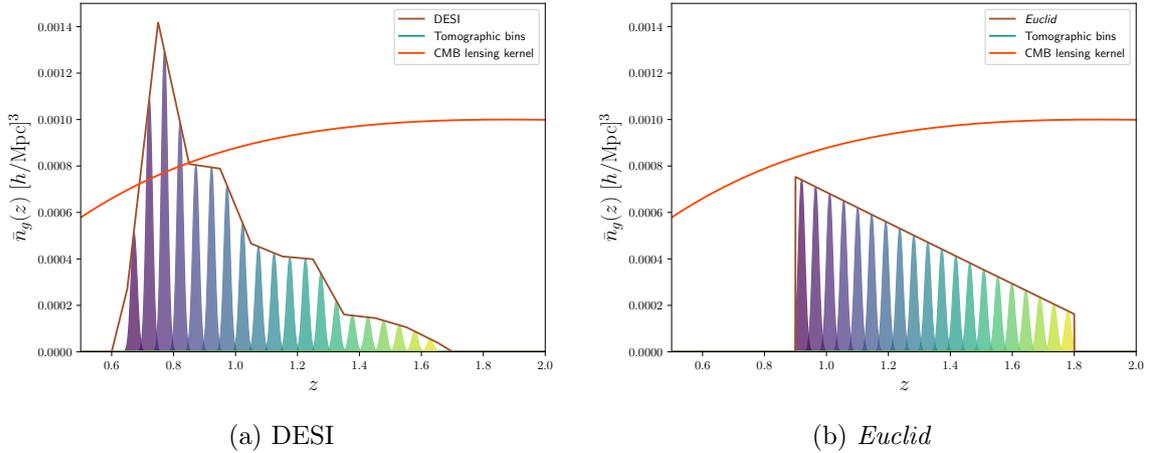


Figure 4.1: Galaxy density distribution as a function of redshift for the emission line galaxies (ELG) of DESI (left panel), and the *Euclid* spectroscopic sample (right panel). The filled coloured lines show the Gaussian bins used in our analysis, colour coded as a function of bin index. The orange line shows the CMB lensing efficiency kernel (with arbitrarily normalization).

assume the model 3 from [Pozzetti et al. \(2016\)](#) for the expected number density of galaxies  $\bar{n}_g(z)$  (see Figure 4.1b). For the expected galaxy linear bias, we fit a linearly redshift dependent bias on the values of the Table 3 of [Euclid Collaboration et al. \(2019\)](#), yielding

$$b_g(z) = 0.79 + 0.68 z . \quad (4.2)$$

### 4.2.3 Tomography

As already mentioned above, our forecasts are based on a tomographic approach where the entire redshift range covered by a galaxy survey is sliced into different redshift bins. Centred at each of these redshift bins, we consider Gaussian redshift shells of a given width  $\sigma_z$  centred on redshifts  $z_i$ ,

$$W_i(z) = \exp\left(-\frac{(z - z_i)^2}{2\sigma_z^2}\right) . \quad (4.3)$$

Provided that a Gaussian shell will be diluting information on radial scales shorter than the Gaussian width, our choice for  $\sigma_z$  is a compromise between maximizing the amount of radial scales under study, and minimizing the impact of non-linear, radial scales in the analysis (Asorey et al. 2012; Di Dio et al. 2014). Hernández-Monteagudo et al. (2019) has shown that, at  $z \simeq 1$ , down to  $\sigma_z = 0.01$  the impact of radial non-linearities was either negligible or easily tractable with a Gaussian kernel describing thermal, stochastic, radial motions. We thus adopt  $\sigma_z = 0.01$  for our forecasts.

As shown in Asorey et al. (2012), the angular galaxy clustering analysis can recover the same amount of information as the 3D analysis when the bin size is comparable to the maximum scale probed by the 3D analysis. This gives  $\sigma_z c/H(z) \simeq 2\pi/k_{\max}$ , so in our case, for  $z = 1$  and  $k_{\max} = 0.2 h \text{ Mpc}^{-1}$  (see Section 4.3), we get  $\sigma_z \simeq 0.01$ , corresponding to our choice of bin size.

For the two galaxy surveys under consideration we take the same number of 20 redshift bins, and since the overlap between consecutive bins is not zero, we account for all cross-correlations between shells in the covariance matrix. In this way, redundant information between different shells is fully accounted for. The redshift bins sample the range from  $z = 0.65$  to  $z = 1.65$  for DESI, and from  $z = 0.9$  to  $z = 1.8$  for *Euclid*. These redshift bins are displayed in Figures 4.1a and 4.1b, together with the expected number density of tracers for each survey.

## 4.3 Observables

We consider three different observables, namely the angular galaxy clustering, the corresponding ARF, and the CMB lensing convergence field. In order to compute the forecasts, we shall restrict to the linear scales, where the cosmological linear theory of perturbations apply. In practice, we ignored all scales above  $k_{\max} = 0.2 h \text{ Mpc}^{-1}$  at all redshifts. This is a conservative approach, as one could consider a scale cutoff which evolves with redshift as in Di Dio et al. (2014). We shall also assume that our observables are Gaussian distributed, and that the information content is completely captured by the two-point momenta, in particular the angular power spectrum, either auto or cross, depending on whether we combine different observables or not.

The basics of the modelling of angular observables have been introduced in Section 3.3. In what follows, we describe in more details the modelling of the angular galaxy density and of the angular redshift fluctuation, and we include the redshift space distortions introduced in Section 1.7.2. The CMB lensing has already been introduced in Section 3.3.2.

### 4.3.1 Galaxy angular density fluctuations

The 3D field of the number density of galaxies is noted as  $n_g(z, \hat{\mathbf{n}})$ , where  $\hat{\mathbf{n}}$  denotes a direction on the sky. The average number density of galaxies at a redshift  $z$  is defined by  $\bar{n}_g(z) = \langle n_g(z, \hat{\mathbf{n}}) \rangle_{\hat{\mathbf{n}}}$ . The 3D field of galaxy density contrast is then given by

$$\delta_g^{3D}(z, \hat{\mathbf{n}}) = \frac{n_g(z, \hat{\mathbf{n}}) - \bar{n}_g(z)}{\bar{n}_g(z)}. \quad (4.4)$$

We assume that the galaxy density contrast traces the dark matter density contrast  $\delta_m$  via a scale-independent bias:  $\delta_g^{3D}(z, \hat{\mathbf{n}}) = b_g(z) \delta_m^{3D}(z, \hat{\mathbf{n}})$ . This bias depends on the properties of the galaxies used as a tracer for each survey, and are given in Eqs. 4.1 and 4.2.

In our analysis, we model the observed redshift of galaxies  $z_{\text{obs}}$  as a 3D field. It is defined as the sum of the redshift induced by the Hubble flow, and the redshift due to the peculiar velocity of galaxies (see Equation 1.47). We neglect other sources of redshift distortions which are significantly smaller than those considered here (Hernández-Monteagudo et al. 2019).

The angular galaxy density field is then modelled by an integral along the line of sight in which, at every redshift  $z$ , only galaxies within the selection function  $W(z_{\text{obs}}; z_i)$  are included:

$$\delta_g^i(\hat{\mathbf{n}}) = \frac{1}{N_g^i} \int_{z=0}^{\infty} dV_{\Omega} \bar{n}_g(z) b_g(z) \delta_m(z, \hat{\mathbf{n}}) W_i[z_{\text{obs}}(z, \hat{\mathbf{n}})], \quad (4.5)$$

where  $N_g^i = \int_{z=0}^{\infty} dV_{\Omega} \bar{n}_g(z) W_i(z)$  is the average number of galaxies per solid angle, under the  $i$ -th selection function  $W_i$  centred on redshift  $z_i$ , and in practice can be computed from an angular average over the survey's footprint.

We next expand the selection function, retaining only linear terms in density and velocity fluctuations, finding

$$\delta_g^i(\hat{\mathbf{n}}) \simeq \frac{1}{N_g^i} \int_{z=0}^{\infty} dV_{\Omega} \bar{n}_g(z) W_i(z) \left[ b_g(z) \delta_m(z, \hat{\mathbf{n}}) + (1+z) \frac{d \ln W_i}{dz} \frac{\mathbf{v}(z, \hat{\mathbf{n}}) \cdot \hat{\mathbf{n}}}{c} \right], \quad (4.6)$$

with the derivative  $d \ln W_i / dz = -(z - z_i) / \sigma_z^2$ .

### 4.3.2 Angular redshift fluctuations

The ARF field represents the spatial variations of the average redshift of galaxies on the sky. The average redshift of galaxies is given by

$$\begin{aligned} \bar{z} &= \frac{1}{N_g^i} \left\langle \int_{z=0}^{\infty} dV_{\Omega} z_{\text{obs}}(z, \hat{\mathbf{n}}) n_g(z, \hat{\mathbf{n}}) W_i[z_{\text{obs}}(z, \hat{\mathbf{n}})] \right\rangle_{\hat{\mathbf{n}}} \\ &= \frac{1}{N_g^i} \int_{z=0}^{\infty} dV_{\Omega} z \bar{n}_g(z) W_i(z). \end{aligned} \quad (4.7)$$

We thus define the ARF field as

$$\delta_z^i(\hat{\mathbf{n}}) = \frac{1}{N_g^i} \int_{z=0}^{\infty} dV_{\Omega} (z_{\text{obs}}(z, \hat{\mathbf{n}}) - \bar{z}) \bar{n}_g(z) [1 + b_g(z) \delta_m(z, \hat{\mathbf{n}})] W_i[z_{\text{obs}}(z, \hat{\mathbf{n}})], \quad (4.8)$$

where we again refer to a redshift bin centred upon  $z_i$ . Expanding the Gaussian selection function at first order and retaining only linear terms in density and velocity,

we find:

$$\delta_z^i(\hat{\mathbf{n}}) \simeq \frac{1}{N_g^i} \int_{z=0}^{\infty} dV_{\Omega} \bar{n}_g(z) W_i(z) \left[ (z - \bar{z}) b_g(z) \delta_m(z, \hat{\mathbf{n}}) + (1+z) \frac{\mathbf{v}(z, \hat{\mathbf{n}}) \cdot \hat{\mathbf{n}}}{c} \left( 1 + (z - \bar{z}) \frac{d \ln W_i}{dz} \right) \right]. \quad (4.9)$$

Note that given the small widths adopted ( $\sigma_z = 0.01$ ), it is safe to assume that the bias  $b(z)$  remains constant within the redshift bin.

### 4.3.3 Angular power spectra

Our statistical tool to test cosmological models are the angular two-point power spectra  $C_{\ell}$  performed over the three fields defined in Sections 4.3.1, 4.3.2 and 3.3.2. Following the formalism introduced in Section 3.3, and assuming that the galaxy bias and the growth factor are scale independent, one can show that our (cross and auto) angular power spectra can be expressed as the convolution of two kernels  $\Delta_{\ell}^A(k)$  and  $\Delta_{\ell}^B(k)$ , corresponding to the fields A and B (see, e.g. [Huterer et al. 2001](#)):

$$C_{\ell}^{A,B} = \frac{2}{\pi} \int dk k^2 P(k) \Delta_{\ell}^A(k) \Delta_{\ell}^B(k), \quad (4.10)$$

where  $P(k)$  is the linear 3D matter power spectrum at  $z = 0$ , function of the wave number  $k$ .

To obtain the theoretical prediction of our angular power spectra, we start from the 2D fields defined in Eqs. 4.6 and 4.9. The velocity field is related to the matter density contrast field via the linearised continuity equation  $\partial \delta_m / \partial t + \nabla \mathbf{v} / a = 0$ . One can show that the angular galaxy clustering kernel is the sum of two terms, one arising from the density of galaxies and the other from the peculiar line of sight velocities,  $\Delta_{\ell}^g = \Delta_{\ell}^g|_{\delta} + \Delta_{\ell}^g|_{v}$  (see e.g. [Padmanabhan et al. 2007](#)):

$$\Delta_{\ell}^{g,i}|_{\delta}(k) = \frac{1}{N_g^i} \int_{z=0}^{\infty} dV_{\Omega} \bar{n}_g(z) W_i(z) b_g(z) D(z) j_{\ell}(k r(z)), \quad (4.11)$$

$$\Delta_{\ell}^{g,i}|_{v}(k) = \frac{1}{N_g^i} \int_{z=0}^{\infty} dV_{\Omega} \bar{n}_g(z) H(z) f(z) D(z) \frac{dW_i}{dz} \frac{j'_{\ell}(k r(z))}{k}, \quad (4.12)$$

where  $j_{\ell}(x)$  is the spherical Bessel function of order  $\ell$  and  $j'_{\ell}(x)$  is its derivative  $j'_{\ell}(x) \equiv dj_{\ell}/dx$ , and  $f(z)$  is the growth rate introduced in Equation 1.49.

One can thus write the power spectrum as the sum of the contributions from the density and from the velocity kernels  $C_{\ell} = C_{\ell}^{\delta,\delta} + 2 C_{\ell}^{\delta,v} + C_{\ell}^{v,v}$ .

The ARF kernel can also be separated into two kernels:

$$\Delta_{\ell}^{z,i}|_{\delta}(k) = \frac{1}{N_g^i} \int_{z=0}^{\infty} dV_{\Omega} \bar{n}_g(z) W_i(z) b_g(z) D(z) (z - \bar{z}) j_{\ell}(k r(z)), \quad (4.13)$$

$$\begin{aligned} \Delta_{\ell}^{z,i}|_v(k) &= \frac{1}{N_g^i} \int_{z=0}^{\infty} dV_{\Omega} \bar{n}_g(z) H(z) f(z) D(z) W_i(z) \\ &\times \left[ 1 + (z - \bar{z}) \frac{d \ln W_i}{dz} \right] \frac{j'_{\ell}(k r(z))}{k}. \end{aligned} \quad (4.14)$$

The kernel function of the CMB lensing convergence field is given by:

$$\Delta_{\ell}^{\kappa}(k) = \frac{3\Omega_{m,0}}{2} \left( \frac{H_0}{c} \right)^2 \int_{r=0}^{r_*} dr \frac{r}{a(r)} \frac{r_* - r}{r_*} D(z(r)) j_{\ell}(k r), \quad (4.15)$$

where  $r_*$  the comoving distance from the observer to the last scattering surface, and  $a$  is the cosmological scale factor.

The top panel of Figure 4.2 shows the angular power spectra of the angular galaxy clustering and ARF, for a Gaussian selection function of width  $\sigma_z = 0.01$  centered on  $z_i = 0.75$  in a DESI-like survey. We show in the same figure the terms arising from the density fluctuation kernel and the peculiar velocity kernel (c.f. Eqs. 4.11 to 4.14). We can see that the peculiar velocity term is relatively more important (compared to the total power spectrum) in the ARF power spectrum than in the angular galaxy clustering power spectrum. To better illustrate this fact, we show in the bottom panel of Figure 4.2 the ratio of the velocity part of the power spectrum (which is the sum  $C_{\ell}^{v,v} + 2C_{\ell}^{\delta,v}$ ) over the total power spectrum for both angular galaxy clustering and ARF. For both fields, the peculiar velocity contribution dominates at low  $\ell$ , while it vanishes to zero for  $\ell > 300$ . At  $\ell = 10$ , the velocity-dependent part in the power spectrum represents around 67% of the total contribution for  $C_{\ell}^{z,z}$ , while it represents only 58% of  $C_{\ell}^{g,g}$ . The difference between the two is even more visible at  $\ell = 60$ , where the velocity contribution represents 55% of  $C_{\ell}^{z,z}$  and only 35% of  $C_{\ell}^{g,g}$ .

This difference is caused by the intrinsic different nature of the angular galaxy clustering and ARF transfer functions: angular galaxy clustering is sensitive to the average of density and velocity under the Gaussian shell, whereas ARF is sensitive to radial derivatives of those fields. For narrow shells, this makes both fields practically uncorrelated (Hernández-Monteagudo et al. 2019), and given the ratio comparison showed in Figure 4.2, one would expect ARF to be more sensitive than angular galaxy clustering to cosmological parameters impacting peculiar velocities.

#### 4.3.4 Numerical recipes

We can rewrite our kernels as

$$\Delta_{\ell}(k) = \int_0^{\infty} dr \mathcal{A}(r) j_{\ell}(k r), \quad (4.16)$$

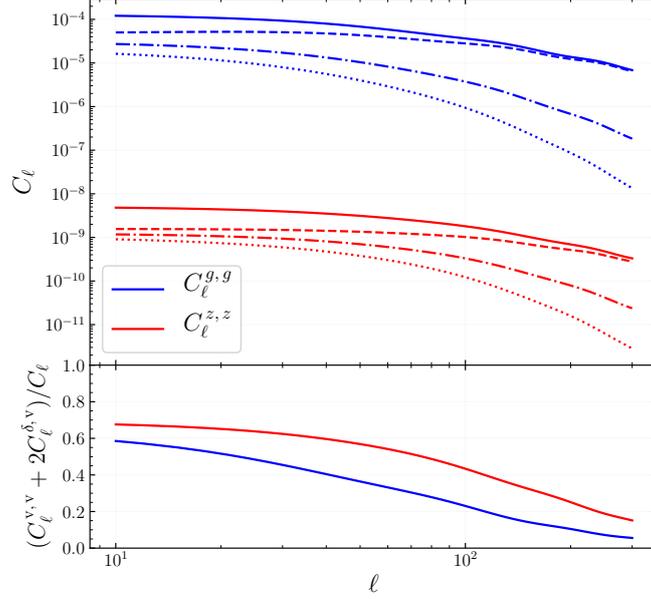


Figure 4.2: *Top panel:* Power spectra of angular galaxy clustering ( $\delta_g$ , in blue) and ARF ( $\delta_z$ , in red), for a Gaussian redshift bin taken in a DESI-like survey, down to  $\ell = 10$ . As described in the text, we do not assume the Limber approximation. The bin is centred on  $z_i = 0.75$  and has a standard deviation of  $\sigma_z = 0.01$ . The dashed line show the term coming from the density kernel  $C_\ell^{\delta,\delta}$ ; the dotted line show the part coming from the velocity kernel  $C_\ell^{v,v}$  and the dot-dashed line show the cross term  $C_\ell^{\delta,v}$ . The total  $C_\ell$  power spectra (plain lines) correspond to the sum  $C_\ell = C_\ell^{\delta,\delta} + 2C_\ell^{\delta,v} + C_\ell^{v,v}$ . *Bottom panel:* Ratio of the velocity dependence in the power spectrum ( $C_\ell^{v,v} + 2C_\ell^{\delta,v}$ ) over the complete power spectrum, for the angular galaxy clustering (blue line) and for the ARF (red line). This figure shows that ARF are more sensitive to the peculiar velocity of galaxies than angular galaxy clustering, for the same redshift shell.

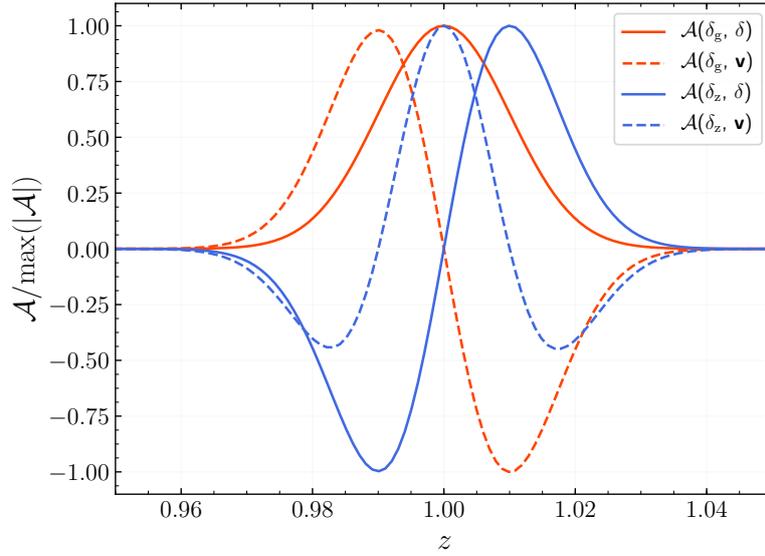


Figure 4.3: Arguments  $\mathcal{A}(z)$  (normalized to their maximum absolute value) of the angular galaxy clustering (blue lines) and of the ARF (red lines), for an *Euclid*-like galaxy survey. The plain lines show the density term, and the dashed lines show the peculiar velocity term. This arguments corresponds to a Gaussian window centred on  $z_i = 1$ . with a standard deviation of  $\sigma_z = 0.01$ .

where we define  $\mathcal{A}(r)$  the argument defined under the integrals of Equations 4.11 to 4.15. For example, the density argument of the  $\delta_g$  observable is

$$\mathcal{A}_\delta^{\delta_g^i}(r) = \frac{1}{N_g^i} r^2 \bar{n}_g(r) W_i(r) b_g(r) D(r). \quad (4.17)$$

We show in Figure 4.3 the arguments for angular galaxy clustering and ARF observables, for both the matter density and from the peculiar velocity terms.

The density argument of the angular galaxy clustering  $\mathcal{A}_\delta^{\delta_g}$  (plain red line) directly reflects our choice of Gaussian selection function. The observable field is indeed the integral along the line of sight of the galaxy 3D overdensity field inside this gaussian selection window. The peculiar velocity argument  $\mathcal{A}_v^{\delta_g}$  (dashed red line) also reflects our choice of Gaussian window, its shape is mainly due to the derivative of our Gaussian selection function  $\frac{dW_i}{dz}$  (see Equation 4.12).

We see that the ARF density term  $\mathcal{A}_\delta^{\delta_z}$  (plain blue line) has a similar shape than  $\mathcal{A}_v^{\delta_g}$ , which seems to be its symmetric around the  $y = 0$  axis. This shape comes from the term  $(z - \bar{z})$  in its argument (see Equation 4.13). The density argument of the ARF field directly probes the evolution in the number of galaxies inside the bin. If there are more galaxies in the  $z < \bar{z}$  half of the bin then the integral along the line of sight will be negative, and if there are more galaxies in the other half of the redshift bin then the integral will be positive. In our equations, the argument is probing the product  $b_g(z) D(z) \bar{n}_g(z)$ . If this product is a constant inside the redshift bin, the integral of the argument along the  $z$ -axis would be zero.

We perform the numerical integration of Equations 4.11 to 4.15 using the FFT-

Log algorithm (Hamilton 2000). This algorithm allows for accurate and fast computations of Hankel transforms, defined by

$$F(k) = \int_0^\infty dr f(r) J_\mu(kr) k, \quad (4.18)$$

with  $J_\mu(kr)$  the Bessel function of order  $\mu$ . The spherical Bessel functions are linked to the Bessel function by

$$j_\ell(x) = \sqrt{\frac{\pi}{2x}} J_{\ell+1/2}(x). \quad (4.19)$$

We rewrite our kernels in order to express them as Hankel transforms

$$\Delta_\ell(k) = \frac{1}{k\sqrt{k}} \int_0^\infty dr \left[ \mathcal{A}(r) \sqrt{\frac{\pi}{2r}} \right] J_{\ell+1/2}(kr) k, \quad (4.20)$$

and we perform the FFTLog algorithm on these rewritten kernels.

We show in Figure 4.4 the resulting kernels obtained for the angular galaxy clustering and the ARF observables, for  $\ell = 10$  and for  $\ell = 300$ . These kernels will be convolved with the matter power spectra as in Equation 4.10 to get the angular power spectra (auto and cross) of our observables. This Figure also validate our choice of maximum scale cutoff, as these kernels are almost negligible for  $k > k_{\max} = 0.2h \text{ Mpc}^{-1}$  for  $\ell = 300$ .

## 4.4 Signal to noise forecasts

We forecast the expected signal to noise ratio (SNR) for different combinations of observables. Our data vector  $\mathbf{D}(\ell)$  contains the auto- and cross-power spectra between the different observables and between the redshift bins. In order to compare several combinations of probes, we define the following data vectors:

$$\mathbf{D}_g(\ell) = (C_l^{g_i, g_j}), \quad (4.21)$$

$$\mathbf{D}_z(\ell) = (C_l^{z_i, z_j}),$$

$$\mathbf{D}_{g,z}(\ell) = (C_l^{g_i, g_j}, C_l^{g_i, z_j}, C_l^{z_i, z_j}),$$

$$\mathbf{D}_{g, \kappa_{\text{CMB}}}(\ell) = (C_l^{g_i, g_j}, C_l^{g_i, \kappa_{\text{CMB}}}, C_l^{\kappa_{\text{CMB}}, \kappa_{\text{CMB}}}),$$

$$\mathbf{D}_{g,z, \kappa_{\text{CMB}}}(\ell) = \left( C_l^{g_i, g_j}, C_l^{g_i, z_j}, C_l^{g_i, \kappa_{\text{CMB}}}, C_l^{z_i, z_j}, C_l^{z_i, \kappa_{\text{CMB}}}, C_l^{\kappa_{\text{CMB}}, \kappa_{\text{CMB}}} \right), \quad (4.22)$$

where  $i$  and  $j$  are indexes running over the redshift bins. We perform a tomographic analysis with 20 redshift bins, thus the data vectors containing only the auto-spectra of angular galaxy clustering and ARF ( $\mathbf{D}_g$  and  $\mathbf{D}_z$ ) contain 210  $C_\ell$ 's each. The data vector containing the cross-correlation  $\mathbf{D}_{g,z}$  has 820  $C_\ell$ 's and the longest data vector  $\mathbf{D}_{g,z, \kappa_{\text{CMB}}}$  contains 861  $C_\ell$ 's.

In Figure 4.5, we display the correlation matrix for the  $\mathbf{D}_{g,z, \kappa_{\text{CMB}}}(\ell = 10)$  data vector. We clearly see that, in the same redshift bin, angular galaxy clustering and ARF are practically un-correlated (diagonal terms of the top left and lower right blocks close to zero), but that there is some degree of anti-correlation in neighbouring

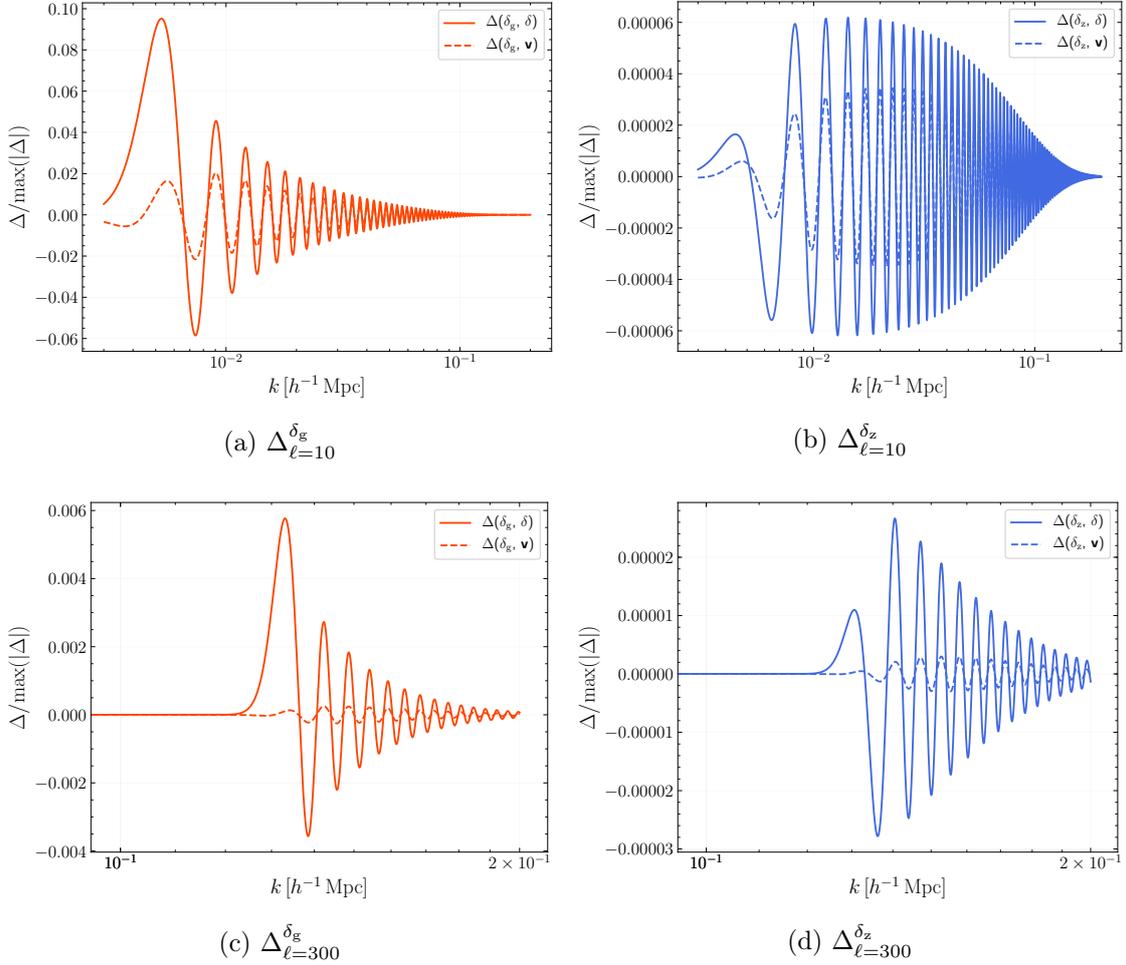


Figure 4.4: The kernels for the angular galaxy clustering and the ARF fields, for a Gaussian bin centred on  $z = 1$  with a width of  $\sigma_z = 0.1$  for an *Euclid*-like survey galaxy distribution. The top row show the kernels for  $\ell = 10$ , and the bottom row show the kernels for  $\ell = 300$ . The left panels are for the  $\delta_g$  observable and the right panels are for the  $\delta_z$  observable. Plain lines show the kernel from the density fluctuations and dashed lines are for the peculiar velocities on the line of sight. These kernels are obtained through the FFTLog algorithm, and are then convolved with the linear power spectrum of matter  $P(k)$  to obtain the angular power spectra  $C_\ell$  of our observables. Note that the axis range differs in every panel.

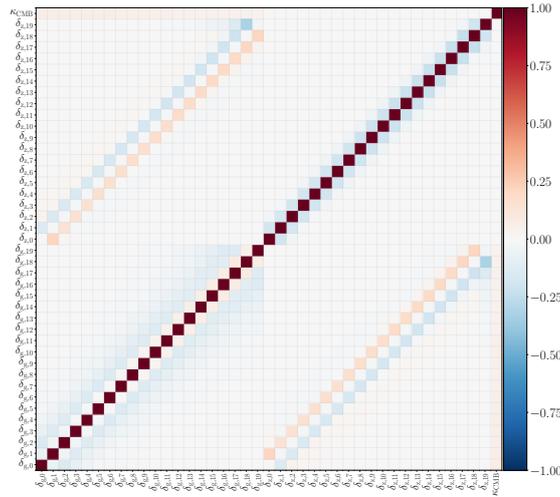


Figure 4.5: Correlation matrix between our observables, for the 20 redshift bins in an *Euclid*-like survey, at  $\ell = 10$ . This matrix corresponds to the  $\mathbf{D}_{g,z,\kappa_{\text{CMB}}}(\ell = 10)$  data vector. The value in each pixel corresponds to  $C_\ell^{\text{A,B}} / \sqrt{C_\ell^{\text{A,A}} C_\ell^{\text{B,B}}}$ . All the data vectors considered in Eqs. 4.21 to 4.22 are a subset of this matrix. We see that there is no correlation between  $\delta_g$  and  $\delta_z$  inside the same redshift bin (diagonals of the upper left and lower right blocks), and that there are opposite and positive correlation for neighboring bins.

redshift bins. We can also observe that the CMB lensing field is almost uncorrelated with the ARF.

The sign of the correlation between adjacent bins can be explained by the shape of the argument functions shown in Figure 4.3. Indeed, the argument function from the density fluctuations in the ARF field is negative and then positive, while the density argument for angular galaxy clustering is always positive. As such, the overlap of the two arguments will be negative if the redshift bin of angular galaxy clustering is centred on a lower redshift than the bin of the ARF, and the correlation will be positive if it is the opposite. As well, correlations between adjacent bins of ARF will be always negative due to the product of the negative and the positive parts of the density argument. Adjacent bins of the angular galaxy clustering have a very low correlation. This is probably due to the fact that the line of sight velocity arguments will be negative and will cancel the positive term coming from the density perturbations. For bins that are further away, the correlation is negative, probably due to the fact that the product of the velocity argument dominates over the product of the density argument.

We assume that there is no correlation between different multipoles and that the covariance between the probes is totally captured by a Gaussian covariance.

The SNR of our data vectors as a function of  $\ell$ , taking into account all redshift bins and the correlations between them, are given by

$$\text{SNR}(\mathbf{D}(\ell)) = \sqrt{\mathbf{D}(\ell)^t \text{Cov}_\ell^{-1} \mathbf{D}(\ell)}, \quad (4.23)$$

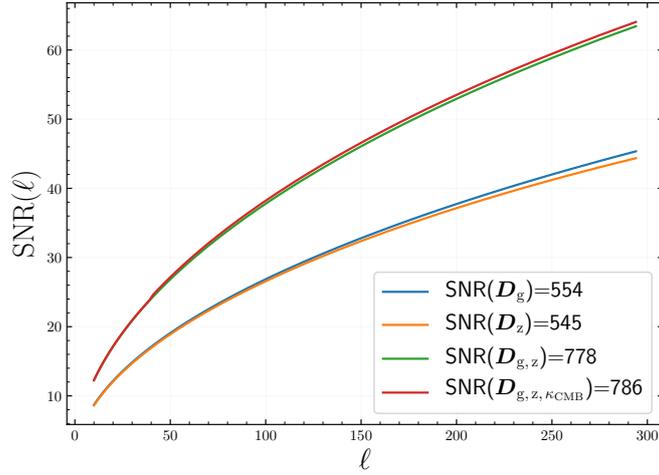


Figure 4.6: Signal to noise ratios of angular galaxy clustering ( $\mathbf{D}_g$ ) in blue, ARF ( $\mathbf{D}_z$ ) in orange and the combinations  $\mathbf{D}_{g,z}$  in green and  $\mathbf{D}_{g,z,\kappa_{\text{CMB}}}$  in red. We use 20 tomographic Gaussian bins of width  $\sigma_z = 0.01$  in an *Euclid*-like survey, in combination with a CMB-S4 survey. The total signal to noise ratios for the range of multipoles  $\ell = 10$  to  $\ell = 300$  are shown in the text box on the bottom right.

and the total SNR are

$$\text{SNR}(\mathbf{D}) = \sqrt{\sum_{\ell=\ell_{\min}}^{\ell_{\max}} [\text{SNR}(\mathbf{D}(\ell))]^2}. \quad (4.24)$$

Assuming that there is no correlation between different multipoles, we define our Gaussian covariance matrix between our data vectors as in [Hu and Jain \(2004\)](#)

$$\begin{aligned} \text{Cov}_\ell \left( C_\ell^{\text{A,B}}, C_\ell^{\text{C,D}} \right) &= \frac{1}{(2\ell + 1) \Delta\ell f_{\text{sky}}} \\ &\times \left[ \left( C_\ell^{\text{A,C}} + \delta_{\text{A,C}}^{\text{K}} N_\ell^{\text{A,A}} \right) \left( C_\ell^{\text{B,D}} + \delta_{\text{B,D}}^{\text{K}} N_\ell^{\text{B,B}} \right) \right. \\ &\left. + \left( C_\ell^{\text{A,D}} + \delta_{\text{A,D}}^{\text{K}} N_\ell^{\text{A,A}} \right) \left( C_\ell^{\text{B,C}} + \delta_{\text{B,C}}^{\text{K}} N_\ell^{\text{B,B}} \right) \right], \end{aligned} \quad (4.25)$$

with A, B, C, D are the observables  $\{g_i, z_j, \kappa_{\text{CMB}}\}$ ,  $\Delta\ell$  is the width of the multipole bin,  $\delta_{x,y}^{\text{K}}$  is the Kronecker delta,  $N_\ell$  are the probe specific noise power spectra and  $f_{\text{sky}}$  is the sky fraction of the survey considered.

For the sake of simplicity, when combining galaxy surveys with CMB lensing, we always assume a full overlap of the two. As such, the sky fraction  $f_{\text{sky}}$  is always taken to be the one of either DESI or *Euclid*. Even if not accurate, this provides a rough estimate of the available constraining power that the combination of galaxy surveys with CMB lensing will be able to achieve.

We assume that the noise of the angular galaxy clustering and that of the ARF are the shot noise arising from the discrete nature of galaxy surveys. We model it by replacing the power spectrum of dark matter by a Poissonian term,  $P_{\text{shot}}(k, z) =$

$1/\bar{n}_g(z)$ , into Eq. 4.10. From this, we can derive the following expressions for the shot noise

$$N_\ell^{g_i, g_j} = \frac{\delta_{i,j}^K}{N_{\text{gal}}^i}, \quad (4.26)$$

$$N_\ell^{z_i, z_j} = \frac{\delta_{i,j}^K}{(N_{\text{gal}}^i)^2} \int dV_\Omega \bar{n}_g(z) W(z_i, z) (z - \bar{z}_i)^2, \quad (4.27)$$

$$N_\ell^{g_i, z_j} = \frac{\delta_j^i}{(N_{\text{gal}}^i)^2} \int dV_\Omega \bar{n}_g(z) W(z_i, z) (z - \bar{z}_i) = 0. \quad (4.28)$$

We can see here that the shot noise cancels out when computing the cross correlation between the angular galaxy density and the ARF fields.

In order to check the validity of the shot noise computations, we performed simulations of the shot noise. We created a uniform random distribution of galaxies, and projected it on a spherical map, based on the same redshift distribution as Euclid. We then perform the computation of the ARF field and of its angular power spectrum on this simulated sky map. Realising several random realisations of this map we obtain an estimation of the shot noise. We found that the theoretical estimation of the shot noise for the ARF field agree at 10% with the numerical estimation with these mock maps.

For the CMB lensing field, we follow the same prescriptions as in the Section 3.5. In particular, we also include the low multipoles ( $\ell < 40$ ) from *Planck* in the forecasts of the Simons Observatory and CMB-Stage 4 surveys.

We use the linear matter power spectrum  $P(k)$  computed with the CLASS software (Blas et al. 2011). In order to focus on the linear regime we restrict our analysis to a maximum multipole of  $\ell_{\text{max}} = 300$ . Assuming the Limber approximation  $k = (\ell + 1/2)/r(z)$ , this  $\ell_{\text{max}}$  corresponds to  $k = 0.18 h \text{ Mpc}^{-1}$  at a redshift of  $z = 0.65$ . Given that we will sample higher redshifts, we will probe larger scales ( $k$  lower than  $0.18 h \text{ Mpc}^{-1}$ ). We hence expect little impact from non-linear physics in our observables (these are expected to become relevant on  $k < 0.2 h \text{ Mpc}^{-1}$  at  $z = 0$ , and yet shorter at higher redshifts). Again this is a conservative approach as one could consider a multipole cutoff evolving with redshift as in Di Dio et al. (2014). We stress that, in our computations, we do not use the Limber approximation but the full computation of spherical Bessel functions.

Our minimum multipole is chosen to be  $\ell_{\text{min}} = 10$ . To reduce numerical noise and to speed up Fisher matrix computations, we perform a linear binning of the multipoles. In each multipole bin  $[\ell_i, \ell_{i+1}[$ , the binned  $C_\ell$  is the average of the  $C_\ell$ 's that fall in the bin, and the binned multipole is taken as  $\ell = (\ell_i + \ell_{i+1})/2$ . We choose a bin size of  $\Delta_\ell = 3$ , which is applied on the full  $\ell$  range. We check that this binning does not impact the constraints from the Fisher matrix by comparing with the case where we do not perform any binning of the multipoles.

We show in Figure 4.6 the SNR for an *Euclid*-like survey combined with a CMB-S4 survey, for four combination of probes:  $\mathbf{D}_g$ ,  $\mathbf{D}_z$ ,  $\mathbf{D}_{g,z}$  and  $\mathbf{D}_{g,z, \kappa_{\text{CMB}}}$  following the redshift binning shown in Figure 4.1b. The total signal to noise for this four data vectors is respectively 544, 545, 778 and 786. This shows that the tomographic

$\Omega_b$	$\Omega_m$	$n_s$	$h$	$\sigma_8$	$w_0$	$w_a$
0.04897	0.3111	0.9665	0.6766	0.8102	-1	0

Table 4.2: Fiducial values of the free parameter of our fiducial cosmological model. We first consider only parameters in the standard  $\Lambda$ CDM model, and later we include the  $w_0$ ,  $w_a$  parameters from the CPL parametrization of Dark Energy.

analysis of angular galaxy clustering and ARF have a similarly high SNR. Moreover, the combined analysis  $\mathbf{D}_{g,z}$  brings more information than measuring the angular galaxy clustering alone  $\mathbf{D}_g$ , as the SNR is increased by 40%.

## 4.5 Fisher forecasts

We use the Fisher formalism which was introduced in Section 3.2.2 to compute *a priori* how well our data vectors defined in Section 4.4 will constrain cosmological parameters in the context of future surveys. As we assumed that there is no correlation between different multipoles, the Fisher matrix can be summed over the multipoles and is given by

$$F_{i,j} = \sum_{\ell_{\min}}^{\ell_{\max}} \frac{\partial \mathbf{D}(\ell)^T}{\partial \lambda_i} \text{Cov}_\ell^{-1} \frac{\partial \mathbf{D}(\ell)}{\partial \lambda_j}, \quad (4.29)$$

with  $\mathbf{D}$  one of the data vectors defined in Eqs. 4.21 to 4.22,  $\{\lambda_i\}_i$  is the set of free parameters of our model, and  $\text{Cov}_\ell$  the covariance matrix given in Eq. 4.25.

The derivatives  $\partial \mathbf{D}(\ell)/\partial \lambda_i$  are computed as the two-point variation with a 1% step around the fiducial value. We have checked that our derivatives are numerically stable when changing the step size.

We compute forecasts for two cosmological models. The first one assumes the standard  $\Lambda$ CDM model, and the parameters we vary are  $\{\Omega_m, \Omega_{\text{baryon}}, \sigma_8, n_s, h, \}$ . The fiducial values of these parameters are given by [Planck Collaboration et al. \(2018a\)](#). The second model assumes an evolving dark energy equation of state, with the so-called CPL parametrization ([Chevallier and Polarski 2001](#); [Linder 2003](#)):  $w(z) = w_0 + w_a z/(1+z)$ . Our second set of free parameters is then

$$\{\Omega_m, \Omega_{\text{baryon}}, \sigma_8, n_s, h, w_0, w_a\}.$$

In both cases, we assume a flat universe ( $\Omega_k = 0$ ) with massless neutrinos ( $\sum m_\nu = 0$ ). We show in Tab. 4.2 the fiducial values of the free parameters.

We also consider a bias parameter assumed constant within each redshift bin, thus adding one free parameter for each redshift shell, over which we marginalise in the Fisher analysis. The fiducial values of the galaxy bias depend on the survey considered and are given in Eqs. 4.1 and 4.2. We take the value at  $z_i$ , the center of the Gaussian shell for each bin.

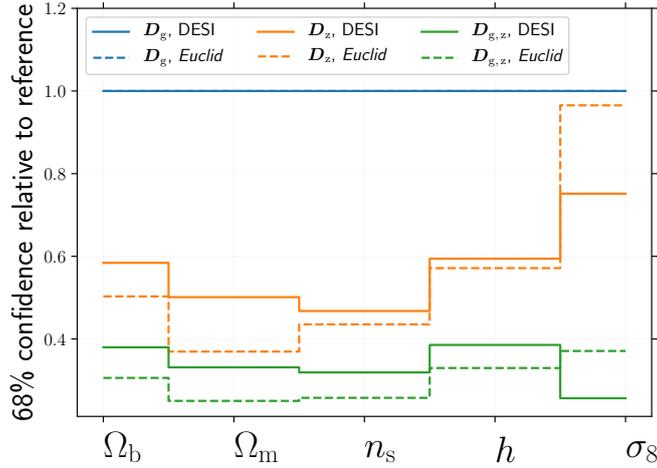


Figure 4.7: Ratio of  $1\sigma$  confidence interval relative to the  $1\sigma$  value from angular galaxy clustering ( $\mathbf{D}_g$ ) for  $\Lambda$ CDM parameters. Constraints are marginalised over the 20 galaxy bias parameters. Plain lines are for a DESI-like survey while dashed lines are for an *Euclid*-like survey. Blue line shows  $\mathbf{D}_g$  (our reference here), orange lines show  $\mathbf{D}_z$  and green lines show  $\mathbf{D}_{g,z}$ . We see that for most parameters (except  $\sigma_8$ ) confidence intervals shrink by  $\sim 50\%$  when using  $\mathbf{D}_z$  instead of  $\mathbf{D}_g$ . When using the combination  $\mathbf{D}_{g,z}$ ,  $1\sigma$  intervals are shrunk by at least 60% for all parameters.

#### 4.5.1 Results for the $\Lambda$ CDM model

The results for the  $\Lambda$ CDM model are summarized in Figure 4.7, where we show the ratio of the  $1\sigma$  marginalised uncertainties when including ARF compared to using only angular galaxy clustering, for a DESI-like and an *Euclid*-like surveys. Figure 4.8 and Figure 4.9 show the  $1\sigma$  uncertainty ellipses for the  $\Lambda$ CDM parameters and three out of the 20 galaxy bias parameters for a DESI-like survey and an *Euclid*-like survey respectively. Error ellipses for  $\mathbf{D}_g$ ,  $\mathbf{D}_z$ , and  $\mathbf{D}_{g,z}$  are given by blue, orange, and green curves, respectively, while marginalised  $1\sigma$  uncertainties for each parameter are quoted, for these three sets of observables, above the panels containing the one dimensional probability density distributions (PDFs).

For both types of LSS surveys, we can see in Figure 4.7 that ARF ( $\mathbf{D}_z$ ) are significantly more sensitive than angular galaxy clustering ( $\mathbf{D}_g$ ), reducing by a factor two the marginalised uncertainties of all cosmological parameters but  $\sigma_8$ , to which both observables are similarly sensitive. For the combined analysis  $\mathbf{D}_{g,z}$ , marginalised uncertainties are reduced by more than 60% for all parameters (including  $\sigma_8$ ), compared to the angular galaxy clustering probe alone  $\mathbf{D}_g$ . We find that using ARF in combination with angular galaxy clustering provides almost the same improvement on the constraints on cosmological parameter for both surveys, although the improvement is on average slightly better for our *Euclid*-like survey.

We see in Figures 4.8 and 4.9 that while the degeneracy direction between different cosmological parameter pairs seems very similar for both angular galaxy clustering and ARF, this is again different for  $\sigma_8$ . For  $\mathbf{D}_z$  this parameter seems rather independent from other cosmological parameters, while its degeneracy with bias pa-

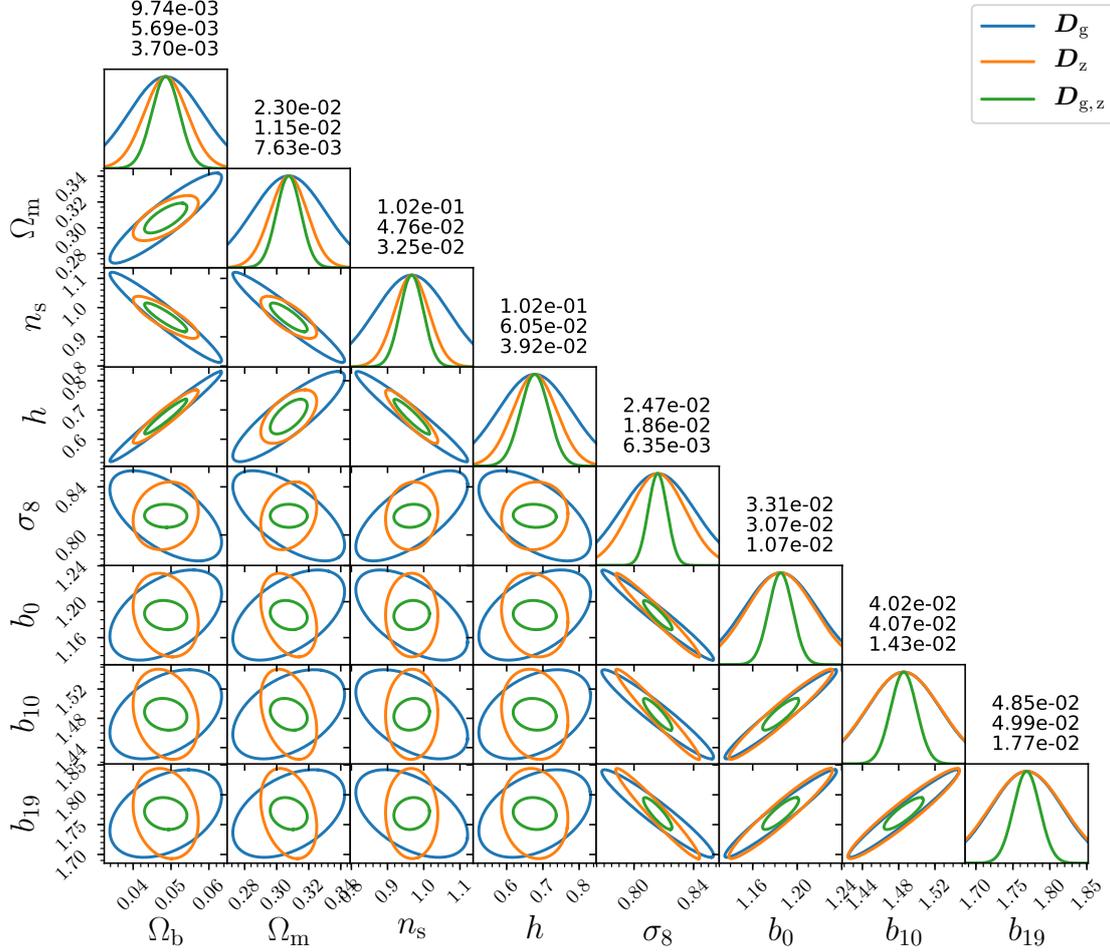


Figure 4.8: Foreseen constraints ( $1\sigma$  contours) for a set of five  $\Lambda$ CDM parameters, plus three galaxy bias parameters (out of a total of 20) for a DESI-like survey. We assume 20 tomographic Gaussian bins of size  $\sigma_z = 0.01$ . The blue line are the constraints for angular galaxy clustering alone  $D_g$ , the orange line are for the ARF alone  $D_z$ , and the green line is a joint analysis of both fields  $D_{g,z}$ . The figures above the 1-D PDFs give the marginalised  $1\sigma$  uncertainty of the parameter for each data vector. We show here only 3 galaxy bias parameters even if we marginalised upon the 20 bias parameters.

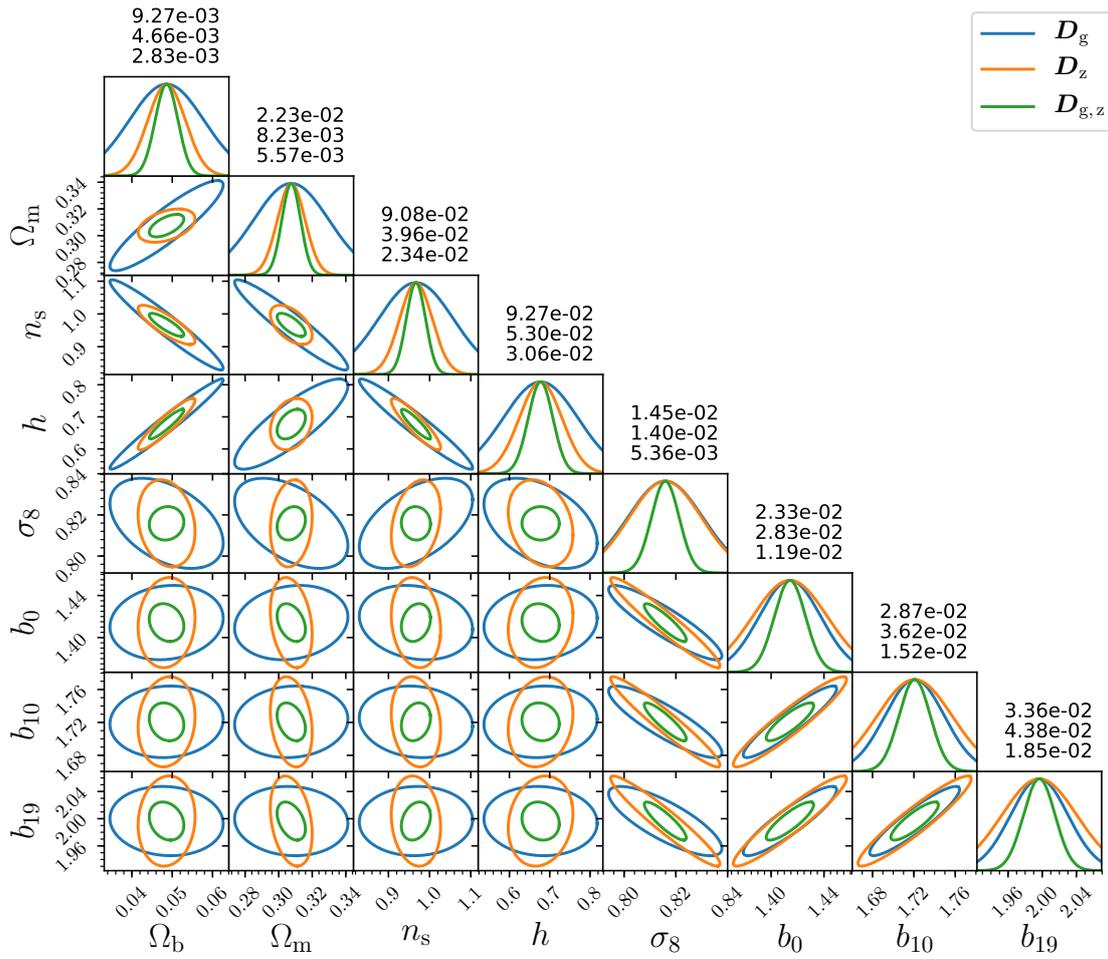


Figure 4.9: Same as Figure 4.8 for an *Euclid*-like survey.

rameters is slightly tilted with respect to that of  $\mathbf{D}_g$ . As a consequence the joint  $\mathbf{D}_{g,z}$  ellipses show little degeneracy with other parameters, including bias. We also find that the marginalised constraints from both experiments are very close, although the *Euclid*-like experiment provides slightly more sensitive forecasts.

The correlation matrices in Figure 4.10 provide an alternative view of our results. It shows the correlation matrices for the 5  $\Lambda$ CDM parameters and the 20 galaxy bias parameters for a DESI-like survey. We see the opposite correlation of the cosmological parameters  $\Omega_b$ ,  $\Omega_m$ ,  $n_s$  and  $h$  with  $\sigma_8$  for angular galaxy clustering and ARF. This opposite correlation is mirrored in the correlations of those three cosmological parameters with galaxy bias parameters. This is expected as  $\sigma_8$  and bias are tightly correlated. The different nature of the correlation of  $\sigma_8$  and bias with the other cosmological parameters for angular galaxy clustering and ARF is critical for (partially) breaking degeneracies when combining angular galaxy clustering with ARF.

Even for those parameters for which both angular galaxy clustering and ARF show a similar direction of degeneracy, the combination of the two observables yields significantly reduced error ellipses. This is mostly due to the lack of correlation between the ARF and angular galaxy clustering under narrow redshift shells, as shown in Hernández-Monteagudo et al. (2019) and in the Figure 4.5.

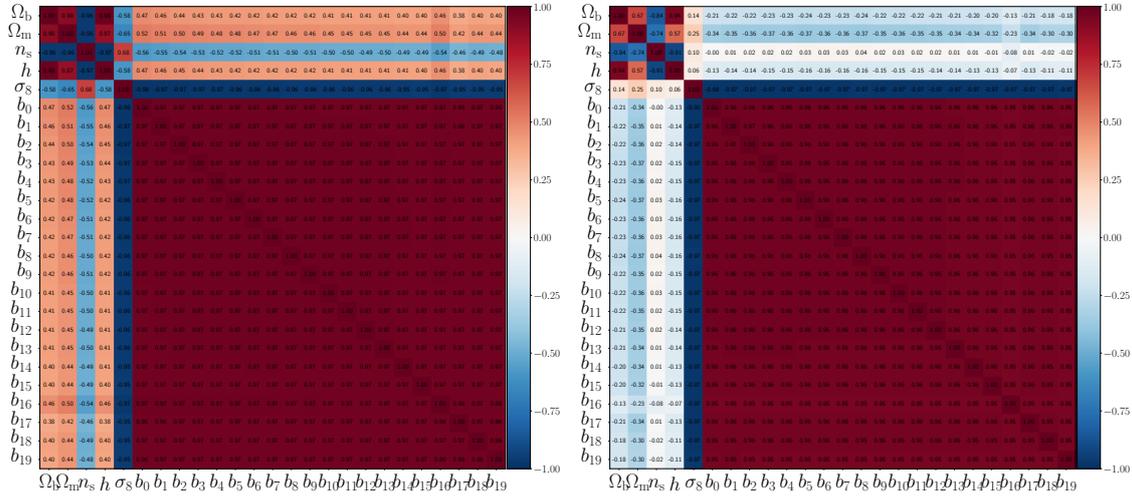
#### 4.5.2 Extension to CPL Dark Energy parametrization

We repeat the analysis detailed above including two new parameters describing the equation of state of dark energy following the CPL parametrization:  $w_0$  and  $w_a$ . We show in Figure 4.11a the improvement on the marginalised uncertainties of the ARF with respect to angular galaxy clustering alone. We find that  $\mathbf{D}_z$  improves the constraints by 20 % and up to 50 % on this set of free wCDM parameters, for both surveys. The combined analysis  $\mathbf{D}_{g,z}$  reduces the uncertainties by at least 50 % and up to 80 % for  $\Omega_m$ ,  $\sigma_8$ ,  $w_0$  and  $w_a$ .

In our idealized case, the combination of ARF with angular galaxy clustering greatly improves the sensitivity of these surveys to dark energy. As shown in Figure 4.11b, the figure of merit of  $w_0$ - $w_a$  increases by more than a factor of 10 when ARF are combined to angular galaxy clustering. It increases from 17 to 189 for our DESI-like survey and from 19 to 345 for our *Euclid*-like survey.

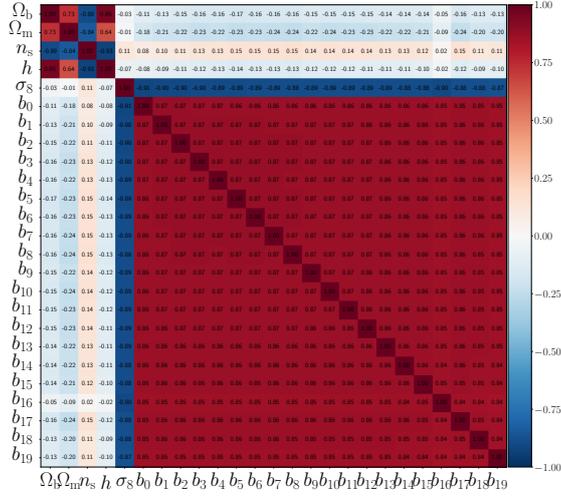
#### 4.5.3 Combining with CMB lensing

We show in Figure 4.12 the improvements on the constraints of the  $\Lambda$ CDM and wCDM parameters for an *Euclid*-like survey, when combined with CMB lensing from *Planck*, Simons Observatory and CMB-S4, marginalised over the galaxy bias parameters. We see that including CMB lensing from *Planck* improves the constraints by maximum of 10 % in both cosmologies. The improvement is more significant when combining with Simons Observatory or CMB-S4. For the Simons Observatory and CMB-S4, in the  $\Lambda$ CDM model, marginalised uncertainties on  $\Omega_m$  and  $\sigma_8$  are decreased by up to 30 %. Other parameters are improved by 5 % to 10 % . For the wCDM model, the improvement is of  $\sim 15$  % for most parameters, with



(a)  $D_g$

(b)  $D_z$



(c)  $D_{g, z}$

Figure 4.10: Correlation between parameters of the  $\Lambda$ CDM model, for a DESI like survey. The top panel is for the angular galaxy clustering alone, the central panel is for ARF alone, and the bottom panel is when combining both observables. We see that the angular galaxy clustering and ARF have opposite correlation coefficients between cosmological parameters and the galaxy bias. The combination of both helps greatly in breaking degeneracies with the galaxy bias.

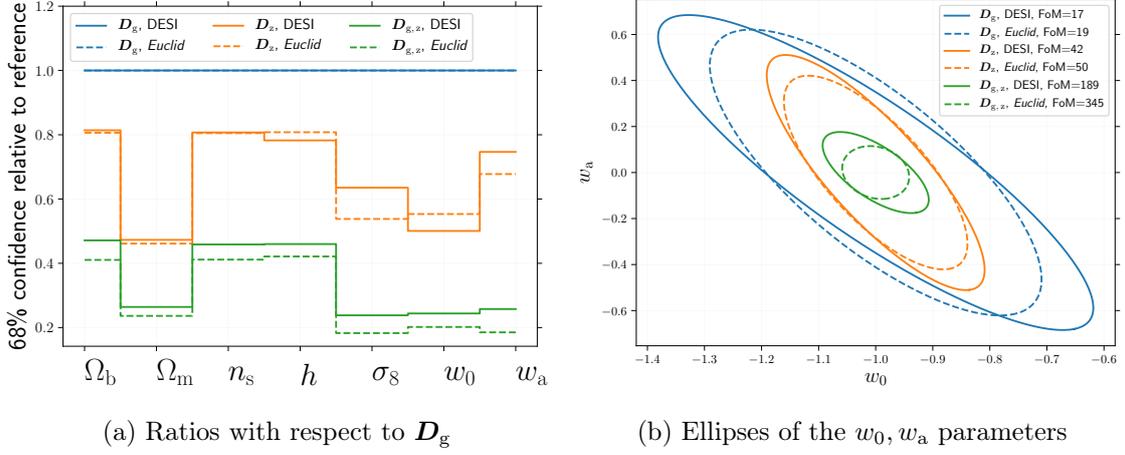


Figure 4.11: *Left*: Ratios of the  $1\sigma$  marginalised uncertainties relative to the  $1\sigma$  marginalised uncertainty for  $D_g$ . We assume  $w$ CDM model and marginalise on 20 galaxy bias parameters (one for each redshift bin). *Right*: Marginalised constraints ( $1\sigma$  contours) on the dark energy equation of state parameters. For both cases, blue lines show  $D_g$ , orange lines show  $D_z$  and green lines show  $D_{g,z}$ . Solid lines are for a DESI-like survey while dashed lines are for an *Euclid*-like survey. The upper right box in the left panel display the figure of merit for each configuration.

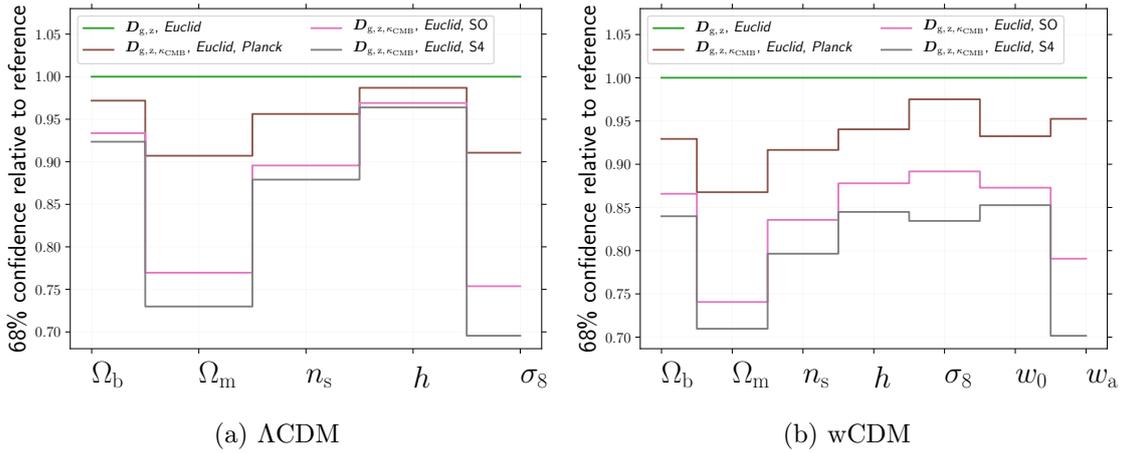


Figure 4.12: Ratio of the  $1\sigma$  constraints in the  $\Lambda$ CDM and  $w$ CDM models (left and right panels respectively), from  $D_{g,z,\kappa_{\text{CMB}}}$  over the  $1\sigma$  constraints from  $D_{g,z}$  for the *Euclid*-like spectroscopic survey. We show combinations with CMB lensing from *Planck* (brown), Simons Observatory (pink) and CMB Stage-4 (grey). Constraints are marginalised over the 20 galaxy bias parameters.

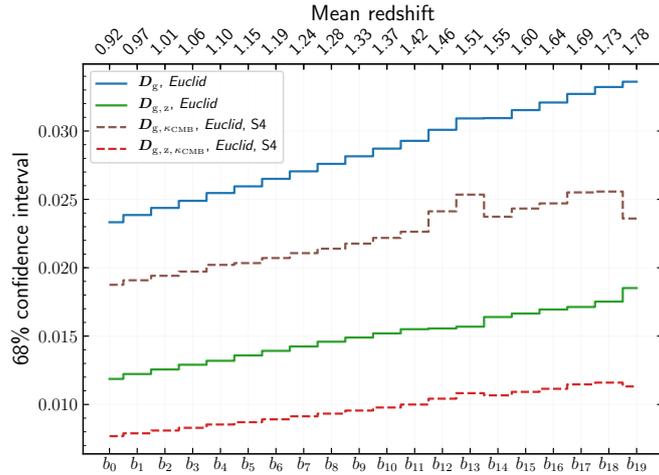


Figure 4.13: Marginalised  $1\sigma$  confidence values for the galaxy bias parameters, with an *Euclid*-like survey alone (plain lines) or in combination with CMB-S4 lensing survey (dashed lines). We marginalised on the 5 free parameters of the  $\Lambda$ CDM model. We show constraints with angular galaxy clustering (blue and brown lines) and in combination with ARF (green and red lines). The mean redshift of each shell is shown at the top. We see that the ARF combined with angular galaxy clustering  $\mathbf{D}_{g,z}$  provides better constraints on galaxy bias than the combination of angular galaxy clustering with CMB lensing  $\mathbf{D}_{g,\kappa_{\text{CMB}}}$ .

the most significant for  $\Omega_m$  and  $w_a$ , with uncertainties decreased by up to 30%. We see that the combination with CMB lensing help decreasing uncertainties on the  $w$ CDM cosmology.

Since the CMB lensing is an unbiased probe of the distribution of matter, one of the main interests of combining it with galaxy surveys is to produce tight constraints on the galaxy bias parameter. We show in Figure 4.13 the  $1\sigma$  marginalised uncertainties on the galaxy bias parameters for each of the 20 redshift bins in an *Euclid*-like survey combined with CMB-S4 lensing, for the  $\Lambda$ CDM model. We compare the constraints obtained for angular galaxy clustering alone ( $\mathbf{D}_g$ ), with the ones obtained when combined with CMB lensing ( $\mathbf{D}_{g,\kappa_{\text{CMB}}}$ ), with ARF ( $\mathbf{D}_{g,z}$ ), and the full combination ( $\mathbf{D}_{g,z,\kappa_{\text{CMB}}}$ ).

We see that the combination of angular galaxy clustering with ARF provides better constraints on the galaxy bias than the combination with CMB lensing. For instance, at a redshift of 1.06, the marginalised uncertainties for the galaxy bias parameter  $b_3$  is of 0.025 for the angular galaxy clustering, it decreases to 0.020 when combined with CMB lensing, and down to 0.013 when combined with ARF. The combination of the three results in marginalised uncertainties of 0.008. We can see that the CMB lensing improves constraints by  $\sim 20\%$  only, while ARF improves constraints by  $\sim 50\%$  (a factor 2 improvement). We argue that this is due to the importance of the velocity term in the ARF kernel (see Fig 4.2), which does not depend on galaxy bias as it is sensitive to the full matter distribution.

## 4.6 Discussion

One could argue several reasons why angular observables might be preferred over standard 3D ones. Probably the main one is the lack of assumption of any fiducial cosmological model to analyse the data. This means that angular observables may be *directly* compared with theoretical predictions without any intermediate data manipulations that hinge on an assumption whose implications in the analysis may not always be clear. Moreover, this type of angular analysis are conducted tomographically in moderately narrow redshift shells, thus avoiding the assumption that the universe remains effectively frozen in relatively long time spans, as it may occur in 3D clustering analysis where an *effective* redshift must be defined for the entire volume under analysis (see, e.g., [Cuesta et al. 2016](#)). [Asorey et al. \(2012\)](#) and [Di Dio et al. \(2014\)](#) have shown that when using a large number of narrow redshift slices, a 2D clustering analysis can produce the same constraints on cosmological parameters than a 3D clustering analysis, provided that the width of the redshift slices is comparable to the minimum scale probed in the 3D analysis. By including the redshift information in a 2D field, the ARF observable keeps some information about the distribution of galaxies along the line of sight, which normally disappear when projecting the 3D galaxy density field on a 2D observable. As we have shown, ARF are improving the usual 2D galaxy clustering analysis.

Another major interest of using angular observables is that they can easily be cross-correlated with other 2D observables. Indeed the combination of 3D probes with 2D probes is not straightforward, especially when one has to properly take into account the covariances between them. In our work we used the CMB lensing field and its cross-correlation with our tomographic analysis of angular galaxy clustering and ARF. We have shown that this cross-correlations improve the constraints, especially on the galaxy bias. [Chaves-Montero et al. \(2019\)](#) have shown that the cross-correlation of the ARF field with the CMB temperature field can detect the kinematic Sunyaev-Zel'dovich (kSZ) effect at the  $10\sigma$  level.

The point of our work is not a detailed comparison between 2D and 3D clustering analysis, but rather an exploration of the added value of including ARF in cosmological studies of the large scale structures, on top of the traditional angular galaxy clustering. By its intrinsic different sensitivity to the cosmic density and velocity fields under the Gaussian redshift shells, the ARF change the degeneracies between cosmological parameters, especially with respect to  $\sigma_8$  and the galaxy bias, compared to the angular galaxy clustering. This is due, as claimed in [Hernández-Monteagudo et al. \(2019\)](#), to the fact that angular galaxy clustering is sensitive to the first moment (the average) of matter density and velocity under the redshift shells, whereas ARF are sensitive to the second moment (the variance) of matter density and velocity along the line of sight, inside these redshift shells. Moreover, we have shown that the ARF and the angular galaxy clustering inside the same tomographic redshift bin are almost uncorrelated. Due to this absence of correlation, by combining both we are able to break degeneracies and give tighter constraints on all the cosmological parameters we have considered.

The results we obtained in our work can be considered as an optimistic setting for both galaxy and CMB surveys. We restricted our analysis to the linear regime

and we did not include any systematic effects which could impact our results and worsen the constraints. It was found in [Hernández-Monteagudo et al. \(2020\)](#) that the impact of non-linear physics is more severe in angular galaxy clustering than in ARF. They found that a linear bias was sufficient to describe the ARF on scales larger than  $60 h^{-1}$  Mpc, while it was not the case for angular galaxy clustering. Indeed, ARF are built upon the average observed redshift along the line of sight in a redshift selection function. This is intrinsically different to counting the number of galaxies in a given region in the universe, and consequently systematics and non-linearities are affecting differently each observable. In future works we plan to address systematics and non-linearities, aiming to model more realistic settings. We expect that the impact of both systematics and non-linearities will depend on the survey and on the targeted galaxy sample, as ongoing work on existing galaxy surveys is indicating.

The point of this work is not a detailed comparison with the forecasted constraints of the *Euclid* survey published in [Euclid Collaboration et al. \(2019\)](#). Indeed our analysis consider a simplistic, linear model of the galaxy clustering. In this context, our findings indicate that ARF brings significant cosmological information on top of the traditional angular galaxy clustering. At best, our results with the angular galaxy clustering probe ( $\mathbf{D}_g$ ) could be compared with the linear setting shown in the Table 9 of [Euclid Collaboration et al. \(2019\)](#) (first line). In that case, their probe is the 3D linear galaxy power spectrum, with a cutoff value at  $k_{\max} = 0.25 h \text{ Mpc}^{-1}$ , in four different redshift bins. Their Fisher analysis account for more parameters describing the anisotropies in the power spectrum and the shot noise residual. This 3D probe is intrinsically different to the (2D) angular power spectrum tomography used in our work, in 20 Gaussian bins, for which we limit to  $k_{\max} = 0.20 h \text{ Mpc}^{-1}$ . Our forecasts with  $\mathbf{D}_g$  for the errors on some parameters are tighter than theirs (by a factor of  $\sim 2$  for  $\sigma_8$ ), while for others we find the opposite situation (e.g., the reduced Hubble parameter  $h$ , whose uncertainty in [Euclid Collaboration et al. 2019](#) is roughly one third of ours).

## 4.7 Conclusion

We showed that the ARF are a promising cosmological observable for next generation spectroscopic surveys. We found that for our choice of binning, the tomographic analysis of ARF retrieves more information than the tomographic analysis of the angular galaxy clustering. We showed that the joint analysis of both fields helps in breaking degeneracies between cosmological parameters, due to their lack of correlation and their different sensitivities to cosmology. The improvement appears to be particularly significant for the  $w_0$ - $w_a$  model. We have shown that the figure of merit for the  $w_0$ - $w_a$  parameters was increased by a factor of more than ten when combining angular galaxy clustering with ARF.

Finally, we have seen that combining angular galaxy clustering with ARF provides tighter constraints on the galaxy bias parameters compared to the combination of angular galaxy clustering with CMB lensing. This shows that ARF are a very powerful probe of the distribution of matter, as it allows to break the degeneracy between  $\sigma_8$  and the galaxy bias. For future galaxy surveys, errors on the cosmologi-

cal figure of merit will be dominated by systematic uncertainties and non-linearities, and ARF might provide a novel, complementary view on those issues.

In our analysis we did not consider massive neutrinos. As the growth rate is particularly sensitive to them, we expect ARF to be a powerful tool to constrain the mass of neutrinos. We defer this detailed analysis to an upcoming work.

Simultaneously from the LSS and CMB fronts, the coincidence in the acquisition of data sets of tremendous quality and huge size should enable the combination of standard analyses with new, alternative ones like the one introduced in this chapter. The combination of techniques and observables should work jointly in the efforts of identifying and mitigating systematics, and pushing our knowledge of cosmological physics to its limits.



# Chapter 5

## Conclusion

During my thesis I investigated the evolution of galaxies and the nature of the Universe. To address these topics, I probed the distribution of matter at different scales. I first focused on the co-evolution of the stellar content, dark matter halo and cold molecular gas to investigate the processes of galaxy formation. I then explored larger scales to forecast the cosmological power of combining upcoming galaxy and CMB surveys.

In the first part of my thesis, I investigated the main drivers of galaxy formation and the interplay between galaxies and their host dark matter haloes. For the first time with a consistent dataset on such a large redshift range, I established the stellar-to-halo mass relation from  $z = 0.2$  to  $z = 4.5$ . I thus confirmed that the stellar-to-halo mass relation evolves with time. Moreover, I tracked the evolution of the characteristic halo mass, which peaks the ratio of stellar mass over halo mass, and found that it increases with redshift up to  $z \sim 2$ , in agreement with other observations, before plateauing up to  $z = 4$ , confirming the results of [Moster et al. \(2013\)](#) and [Behroozi et al. \(2018\)](#). Among the scenarios I discussed to explain the redshift evolution of the characteristic halo mass, I focused on cold gas inflows. To do so, I carried out an innovative approach by combining observations of the cold molecular gas content of galaxies with the stellar-to-halo mass relation. Based on these original results, I found that the evolution of the halo mass of galaxies with a constant cold gas to stellar mass ratio increases from  $z = 0$  to  $z \sim 2$  before plateauing up to  $z \sim 4$ , displaying an evolution similar to the evolution of the characteristic halo mass. This confirms that the cold molecular gas is one of the major drivers of the high efficiency of the galaxy formation inside massive haloes ( $M_h > 10^{12} M_\odot$ ) at  $z \gtrsim 2$ . Observations of the cold gas filaments in galaxy clusters at  $z \gtrsim 2$ , such as those of [Salomé et al. \(2006\)](#) in the local Universe, would be key to better understand the relation between the distribution of the cold molecular gas and galaxy formation across time.

In the second part of my thesis work, I focused on the large scale structure of the Universe in order to understand its nature and its evolution. Specifically, I investigated how next generation wide-field surveys will improve our understanding of the dark sector. As a member of the *Euclid* consortium and its CMBX science working group, I have contributed to the various analyses aiming at probing the gain of joint analyses of the *Euclid* survey with current and upcoming CMB obser-

vations: *Planck*, the Simons Observatory and the CMB-Stage 4. My participation to this collaborative work has covered all the aspects of the project, with specific contributions including the modelling of the experimental noise, the tests and implementation of the Fisher analysis and the writing of the article. In this work, we showed that combining *Euclid* with CMB surveys will greatly improve most cosmological parameters, help in breaking degeneracies and in lowering the importance of systematic uncertainties. In this idealistic modelling, we also found that the inclusion of the cross-correlations between *Euclid* and CMB observables brings marginal improvements. A more careful modelling of the systematic effects affecting each experiment is needed to assess the full potential of the cross-correlations.

Finally, in the scope of maximizing the cosmological power of upcoming spectroscopic surveys, I show how a novel observable, the angular redshift fluctuations (ARF), complements the standard angular galaxy clustering. By performing for the first time a realistic forecast with this new observable, I found that a tomographic analysis combining ARF and angular galaxy clustering will break degeneracies between cosmological parameters. I also discussed how this angular probe can be complementary to traditional 3D probes for spectroscopic surveys. In addition, I found that a joint analysis with the CMB lensing observable further improves constraints on the galaxy bias parameters. Most notably, I demonstrated for the first time that the use of ARF will greatly improve the constraints on the nature of the dark energy. Furthermore, its expected power on other aspects of the dark sector, such as massive neutrinos, and modified gravity, is very promising and will be tested in future works which I plan to contribute to.

Next generation galaxy surveys such as *Euclid*, DESI, J-PASS, or LSST will offer us a huge volume of the Universe to observe, with a high resolution up to the very small scales. We absolutely need to fully exploit these small scales up to  $k \sim 10h \text{ Mpc}^{-1}$ , as they contain a lot of cosmological information, which was even named the small scale miracle by [Lacasa \(2019\)](#). However, we know these scales are where baryonic effects come into play and plague our modelling of the power spectrum (see e.g. [Rudd et al. 2008](#); [Chisari et al. 2018](#)).

Understanding these small scales is one of the major challenges that we will need to face. In the coming years, I will contribute to the endeavor of including astrophysical information such as the stellar mass, star-formation rate and galaxy environments, into cosmological analyses, knowing that these astrophysical properties will be measured, as part of the legacy science, by upcoming surveys. A first step would be for example to develop a model such as the baryonic correction model of [Schneider and Teyssier \(2015\)](#), based on a modification of the halo model, and to test its validity on hydrodynamical simulations. For this, I will benefit from my dual expertise, acquired during my thesis, in the formation and evolution of galaxies and in the cosmological analysis of galaxy surveys. By a careful modelling of the interplay between galaxies, baryons and dark matter, I will contribute to the construction of a state-of-the-art modelling of the distribution of galaxies at small scales. This absolutely crucial step will allow us to fully unleash the power of next generation surveys, and optimise their scientific exploitation to which I will contribute.

In addition, I will pursue my work on the combination of CMB and galaxy surveys with a focus on the CMB lensing as another probe of the small-scales distribution

---

of matter. I propose to calibrate the bias between galaxies and the dark matter density field using the CMB lensing field. To do so, we will need to go beyond the standard estimations of the CMB lensing field, known to be noisy at small scales. As a matter of fact, the upcoming CMB surveys such as Simons Observatory and CMB-S4 will drastically improve the resolution on the small scales, going down to one arcminute. I will develop new estimators of the CMB lensing field targeting the small scales, following for example the works of [Hadzhiyska et al. \(2019\)](#) and [Schaan and Ferraro \(2019\)](#). In the future, I will explore all possibilities offered to me to participate in these future CMB surveys in order to take advantage of their unique information on the dark matter distribution, and combine it with the information on the galaxy distribution provided by *Euclid*.

It is only thanks to these optimised combinations of information that we will be able to unveil some of the mysteries of the dark sector of the Universe.



# Appendix A

## CMB lensing

This appendix introduces a brief summary of the weak lensing of the CMB. A more detailed description can be found in the review [Lewis and Challinor \(2006\)](#) and in the book [Dodelson \(2017\)](#).

### A.1 Introduction

Under the effect of gravity, a light ray will be deflected by the mass it meets along its path, as illustrated in [Figure A.1](#). For a point source, and in the thin lens approximation, the deflection angle  $\alpha$  between the perturbed and original line of sight is governed by the mass of the lens  $M$  and by the impact parameter  $\xi$  (the shortest distance between the lens and the light ray)

$$\delta\theta = \frac{4MG}{c^2 \xi} \quad (\text{A.1})$$

The CMB lensing corresponds to a displacement of the temperature field. The temperature at the position  $\mathbf{n}$  will be seen at the position  $\mathbf{n} - \boldsymbol{\alpha}$

$$T(\mathbf{n}) = T^u(\mathbf{n} - \boldsymbol{\alpha}) \quad (\text{A.2})$$

with  $T$  the observed (lensed) temperature field and  $T^u$  the unlensed temperature field. We can decompose the deflection vectorial field into a potential and rotational term. We neglect the rotational term and we introduce the lensing potential  $\phi$  given by  $\boldsymbol{\alpha} = \nabla\phi$ . The convergence field  $\kappa$  is defined by

$$\nabla^2\phi = -2\frac{\Sigma}{\Sigma_{cr}} = -2\kappa \quad (\text{A.3})$$

with  $\Sigma = \int \rho dz$  the surface mass density, and

$$\Sigma_{cr} = \frac{c^2}{4\pi G} \frac{D_S}{D_L D_{LS}}, \quad (\text{A.4})$$

the critical surface mass density, defined using the angular diameter distance  $D_S$  between the observer and the source,  $D_L$  between the observer and the lens, and

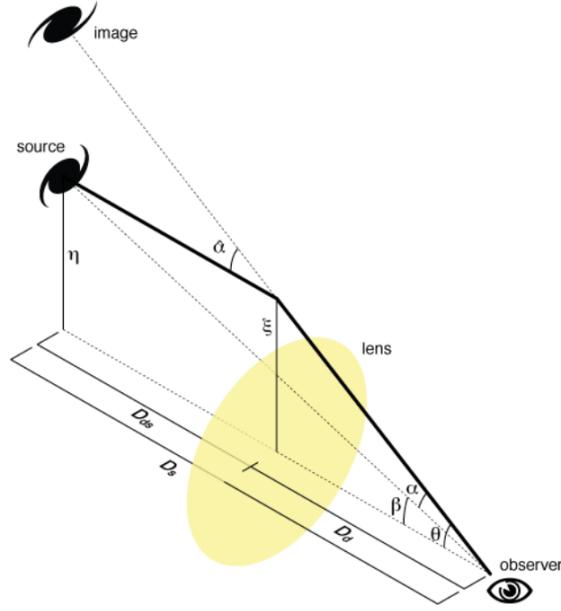


Figure A.1: Deflection of a light ray coming from a point source by a mass present on the line of sight.

$D_{LS}$  between the lens and the source.

## A.2 Statistical anisotropy

The unlensed CMB temperature field is statistically isotropic, i.e. there is no correlation between different multipoles. As we show below, the lensing of the CMB will create statistical anisotropies. The power spectrum of the CMB will not be the same in each direction, depending if there is a mass on this line of sight or not.

If  $\alpha$  remains small, we can do a Taylor expansion of the observed temperature in Equation A.2

$$T(\mathbf{n}) = T^u(\mathbf{n}) - \nabla T^u(\mathbf{n}) \nabla \phi + O(|\nabla \phi|^2) \quad (\text{A.5})$$

if we limit ourselves at the first order in  $\nabla \phi$ . We go to the Fourier space (denoted by hats  $\hat{T}$ ) and we have

$$\hat{T}(\boldsymbol{\ell}) \simeq \hat{T}^u(\boldsymbol{\ell}) + \int \frac{d\ell_1^2}{(2\pi)^2} \boldsymbol{\ell}_1(\boldsymbol{\ell} - \boldsymbol{\ell}_1) \hat{T}^u(\boldsymbol{\ell}_1) \hat{\phi}(\boldsymbol{\ell} - \boldsymbol{\ell}_1). \quad (\text{A.6})$$

where we have assumed the flat sky approximation, i.e. we identify the Fourier modes to the spherical harmonics.

With this we can compute the non diagonal terms of the covariance matrix of the observed temperature field

$$\langle \hat{T}(\boldsymbol{\ell})(\hat{T}(\boldsymbol{\ell}'))^* \rangle_{\boldsymbol{\ell} \neq \boldsymbol{\ell}'} \simeq \int \frac{d\ell_1^2}{(2\pi)^2} \boldsymbol{\ell}_1(\boldsymbol{\ell} - \boldsymbol{\ell}_1) \hat{\phi}(\boldsymbol{\ell} - \boldsymbol{\ell}_1) \langle \hat{T}^u(\boldsymbol{\ell}_1)(\hat{T}^u(\boldsymbol{\ell}'))^* \rangle + c.c., \quad (\text{A.7})$$

where c.c. stands for complex conjugate of the integral, and the brackets mean averaging over several realisations of the unlensed CMB for a fixed lensing potential.

Because the unlensed CMB is statistically isotropic, we have  $C_\ell$  the unlensed power spectrum defined by

$$\langle \hat{T}^u(\boldsymbol{\ell})(\hat{T}^u(\boldsymbol{\ell}'))^* \rangle = (2\pi)^2 \delta_D(\boldsymbol{\ell} - \boldsymbol{\ell}') C_\ell, \quad (\text{A.8})$$

with  $\delta_D$  the Dirac function. Replacing the average in the integral of Equation A.7 by the Delta function, we obtain the relation between the observed correlations of  $T$ , the unlensed  $C_\ell$  and  $\phi$

$$\langle \hat{T}(\boldsymbol{\ell})(\hat{T}(\boldsymbol{\ell}'))^* \rangle_{\boldsymbol{\ell} \neq \boldsymbol{\ell}'} \simeq \boldsymbol{\ell}'(\boldsymbol{\ell} - \boldsymbol{\ell}') \hat{\phi}(\boldsymbol{\ell} - \boldsymbol{\ell}') C_\ell. \quad (\text{A.9})$$

We see here that indeed the CMB lensing potential  $\phi$  will create statistical anisotropies, i.e. non diagonal correlations in the CMB temperature field.

### A.3 Quadratic estimator

We only have one realisation of the CMB, so we can build a really simple estimator  $\tilde{\phi}$  by taking  $\mathbf{L} = \boldsymbol{\ell} - \boldsymbol{\ell}'$ , and we have for any given  $\mathbf{L}$  and  $\boldsymbol{\ell}'$

$$\tilde{\phi}(\mathbf{L}) = \frac{\hat{T}(\boldsymbol{\ell}) \hat{T}^*(\boldsymbol{\ell} - \mathbf{L})}{C_{|\boldsymbol{\ell} - \mathbf{L}|} \mathbf{L}(\boldsymbol{\ell} - \mathbf{L})}. \quad (\text{A.10})$$

This estimator combines two Fourier modes (hence ‘‘quadratic estimator’’), each with a wave number close to one another and separated by the vector  $\mathbf{L}$ .

It is not optimal as it uses only one of all the scales  $\boldsymbol{\ell}$  we can measure in the temperature map. A way to combine all scales is to use the optimized weight functions given by Okamoto and Hu (2003), which weights the  $\boldsymbol{\ell}$  modes in order to optimally to recover the lensing power spectrum.

$$\tilde{\phi}(\mathbf{L}) = N(\mathbf{L}) \int \frac{d^2\ell}{2\pi} \hat{T}(\boldsymbol{\ell}) \hat{T}^*(\boldsymbol{\ell} - \mathbf{L}) g(\boldsymbol{\ell}, \mathbf{L}), \quad (\text{A.11})$$

with  $g(\boldsymbol{\ell}, \mathbf{L})$  the weight function, and  $N(\mathbf{L})$  the normalisation given by

$$N(\mathbf{L})^{-1} = \int \frac{d^2\ell}{(2\pi)^2} \frac{\left[ (\mathbf{L} - \boldsymbol{\ell}) \cdot \mathbf{L} C_{|\boldsymbol{\ell} - \mathbf{L}|}^u + \boldsymbol{\ell} \cdot \mathbf{L} C_\ell^u \right]^2}{2C_\ell^{\text{tot}} C_{|\boldsymbol{\ell} - \mathbf{L}|}^{\text{tot}}}, \quad (\text{A.12})$$

with  $C_\ell^u$  the unlensed CMB temperature power spectrum, and  $C_\ell^{\text{tot}}$  the lensed power spectrum including the noise. This normalisation is chosen so that the estimator is unbiased at the lowest order, i.e if we could average on several CMB realisations we should obtain the true lensing potential  $\phi$ .

In fact, this estimator will be inherently noisy due to the presence of statistical (gaussian) fluctuations of the CMB. As we cannot average over several CMB realisations this noise will contaminate our estimator. This noise can be estimated in the power spectrum of  $C_L^{\phi, \phi}$ , and is often called the N0 bias. It appears that it is

equal to the normalisation defined above in Equation [A.12](#).

We showed here only the impact of lensing on the temperature field. In fact the lensing also affects the polarisation field. In practice it will convert E modes into B modes. We can also define a quadratic estimator on this E and B fields, as well as quadratic estimators on the cross-correlations between this three fields. The minimum variance (MV) estimator is obtained by combining the estimators from the auto- and cross-correlations of the different fields via an inverse variance weighting, in order to keep only the less noisy estimators.

# Appendix B

## Impact of massive neutrinos

Neutrinos are massless in the Standard Model of particle physics. There are three flavours of neutrinos: electronic, muonique and tauique. Each neutrino flavour is a linear combination of three states with a definite mass. There are evidences that neutrinos oscillate between these three flavours thanks to solar, atmospheric, reactor and accelerator observations (e.g. [Fukuda et al. 1998](#)). This oscillation implies that neutrinos have a non zero mass. Oscillation experiments are able to measure the square of the difference of masses, but not the actual mass of each state nor their ranking. Figure [B.1](#) present the two possible mass configurations: normal hierarchy if the smaller mass interval is between the two less massive neutrinos, and inverted hierarchy if the smaller mass interval is between the two most massive neutrinos. The neutrino density is given by

$$\Omega_\nu = \frac{M_\nu}{93.14h^2\text{eV}}, \quad (\text{B.1})$$

and the current bound on neutrino mass is

$$0.06\text{eV} \lesssim M_\nu \equiv \sum_i m_{\nu,i} \lesssim 0.14\text{eV}, \quad (\text{B.2})$$

with  $m_{\nu,i}$  the mass of each neutrino species. The lower bound corresponds to the normal hierarchy and is given by oscillation experiments ([Gonzalez-Garcia et al. 2012](#)). The upper bounds comes from CMB and large scale structures analysis ([Costanzi et al. 2014](#)), because as we will show in this section, massive neutrinos have an impact on the growth of structure.

After the big-bang, neutrinos decouple from matter when the universe cools down, i.e. when the interaction rate of electrons and neutrinos is lower than the expansion rate of the universe, in a process similar to the decoupling of photons and matter. However neutrinos decouple from matter much earlier than photons as they have a much lower interaction rate. Neutrinos behave like photons during the radiation and matter dominated era. One can take into account their impact on the energy budget by adding an *effective number* of relativistic species  $N_{\text{eff}}$  to the

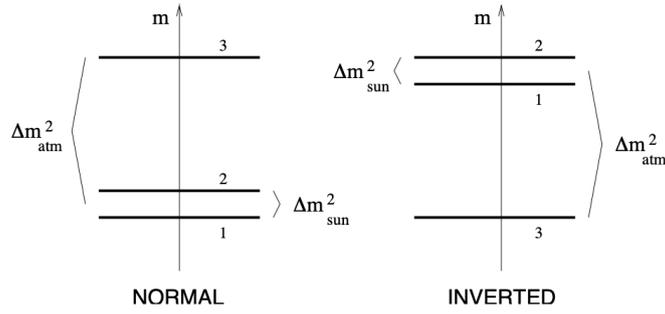


Figure B.1: Figure from [Lesgourgues and Pastor \(2006\)](#). The two possible neutrino schemes: normal hierarchy (NH) and inverted hierarchy (IH).

radiation density ([Shvartsman 1969](#); [Steigman et al. 1977](#))

$$\rho_R = \left[ 1 + \frac{7}{8} \frac{4^{4/3}}{11} N_{\text{eff}} \right] \rho_\gamma, \quad (\text{B.3})$$

In an universe with three relativistic neutrino species we have  $N_{\text{eff}} = 3.046$ . Any departure from this number is an hint of non-standard neutrinos or of the presence of another relativistic species ([Dolgov 2002](#)).

The free-streaming scale is the scale under which a particle cannot be contained in a gravitational potential well due to its dispersion velocity. For scales larger than the free-streaming scale, neutrinos cannot escape the potential well and cluster together with baryons and dark matter. However, for scales smaller than this free-streaming scale, neutrinos will suppress the growth of matter perturbations ([Lesgourgues and Pastor 2006](#); [Levi and Vlah 2016](#)).

Massive neutrinos become non-relativistic when their mass become non-negligible with respect to their total energy. This evolution is shown on the [Figure B.2](#). At this moment they are slowed down and their velocity dispersion will not allow them to escape from large gravitational potential wells. The free streaming scale of neutrinos reaches a maximum. This maximum scale is noted  $k_{\text{nr}}$ . The impact of neutrino is characterized by the parameter  $f_\nu \equiv \Omega_\nu / \Omega_m$ , which is constant once neutrinos become non-relativistic.

Massive neutrinos thus have a scale dependant effect on the growth of structures. In [Figure B.3](#) we see that massive neutrinos will impact the power spectrum for scales  $k > k_{\text{nr}} \sim 10^{-2} h \text{ Mpc}^{-1}$ , by decreasing the amplitude of fluctuations by at least 4%. However on large scales, for  $k < k_{\text{nr}}$ , neutrinos cluster similarly to dark matter and baryons and do not change the power spectrum. The impact on the matter power spectrum for scales  $k \gg k_{\text{nr}}$  is often approximated by ([Hu et al. 1998](#))

$$\frac{P^{f_\nu}(k)}{P^{f_\nu=0}(k)} \sim 1 - 8f_\nu. \quad (\text{B.4})$$

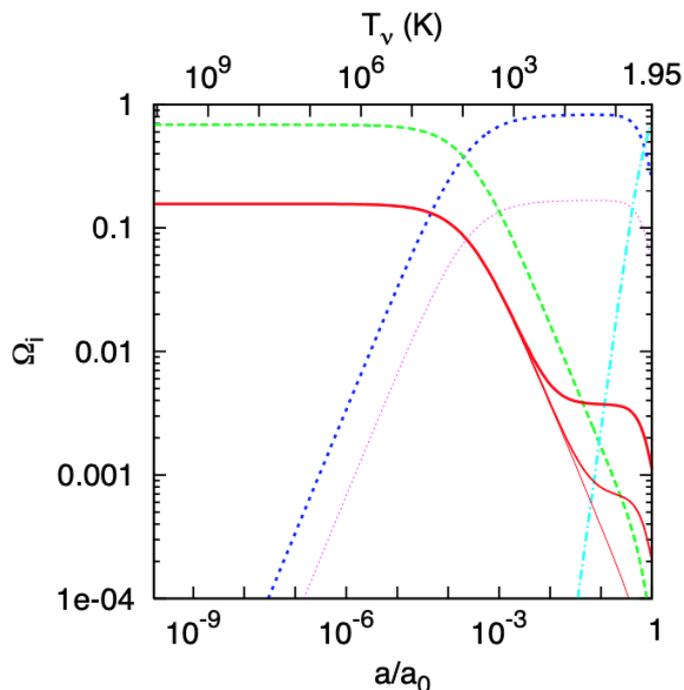


Figure B.2: Figure from [Lesgourgues and Pastor \(2006\)](#). Evolution of the background densities from soon after the time of neutrino decoupling until now. The three neutrino masses are distributed according to the Normal Hierarchy scheme with  $m_1 = 0$ ,  $m_2 = 0.009$  eV and  $m_3 = 0.05$  eV. The top axis shows the neutrino temperature. The decoupling from photons and matter happens when  $a/a_0 \sim 10^{-3}$ . The density of the neutrino mass states  $\nu_2$  and  $\nu_3$  is clearly enhanced once they become non-relativistic, when  $a/a_0 \sim 10^{-2}$ .

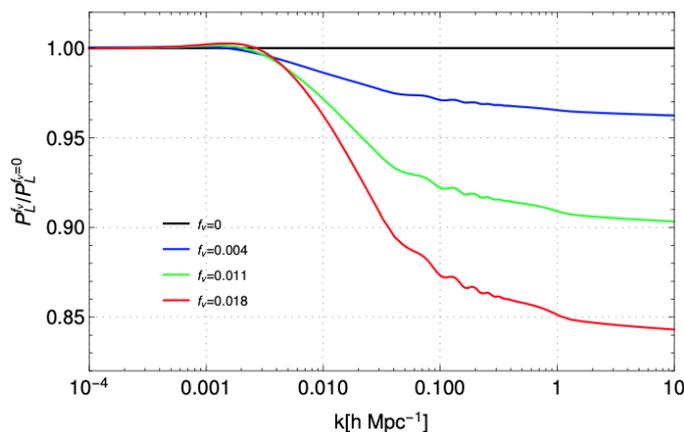


Figure B.3: Figure from [Levi and Vlah \(2016\)](#). Ratio of the matter power spectrum at  $z = 0$  with three massive neutrinos over one with three massless neutrinos. The solid lines correspond to total neutrino mass values  $M_\nu = 0, 0.06, 0.15, 0.24$  eV, parametrized by the neutrino fraction  $f_\nu = 0, 0.004, 0.011, 0.018$ , respectively. The baryon and total matter densities are fixed, and the neutrino and CDM densities are varied accordingly. The non-relativistic transition is located at  $k_{\text{nr}} \sim 10^{-3} - 10^{-2} h \text{ Mpc}^{-1}$ .



# Appendix C

## Related publications

In this appendix I include the publications related to the work presented in my thesis. The first paper [Legrand et al. \(2019\)](#) corresponds to the Chapter 2 and has been published in MNRAS. The second paper [Legrand et al. \(2020\)](#) corresponds to the Chapter 4 and has been submitted to A&A. Finally, the collaborative work of [Euclid Collaboration \(in prep.\)](#) presented in Chapter 3 has not yet been sent for review to the Euclid consortium so I do not include it in this Appendix.

# The COSMOS-UltraVISTA stellar-to-halo mass relationship: new insights on galaxy formation efficiency out to $z \sim 5$

L. Legrand<sup>1,2\*</sup>, H. J. McCracken<sup>2</sup>, I. Davidzon<sup>3</sup>, O. Ilbert<sup>4</sup>, J. Coupon<sup>5</sup>, N. Aghanim<sup>1</sup>, M. Douspis<sup>1</sup>, P. L. Capak<sup>3</sup>, O. Le Fèvre<sup>4</sup> and B. Milvang-Jensen<sup>6,7</sup>

<sup>1</sup>*Institut d'Astrophysique Spatiale, CNRS (UMR 8617), Université Paris-Sud, Bâtiment 121, 91405 Orsay Cedex, France*

<sup>2</sup>*Institut d'Astrophysique de Paris, Sorbonne Universités, UPMC Univ Paris 6 et CNRS, UMR 7095, 98 bis bd Arago, F-75014 Paris, France*

<sup>3</sup>*IPAC, Mail Code 314-6, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA*

<sup>4</sup>*Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France*

<sup>5</sup>*Department of Astronomy, University of Geneva, Ch. d'Ecogia 16, CH-1290 Versoix, Switzerland*

<sup>6</sup>*Cosmic Dawn Center (DAWN), Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen Ø, Denmark*

<sup>7</sup>*DTU-Space, Technical University of Denmark, Elektrovej 327, DK-2800 Kongens Lyngby, Denmark*

Accepted 2019 April 24. Received 2019 April 24; in original form 2018 October 23

## ABSTRACT

Using precise galaxy stellar mass function measurements in the COSMOS field we determine the stellar-to-halo mass relationship (SHMR) using a parametric abundance matching technique. The unique combination of size and highly complete stellar mass estimates in COSMOS allows us to determine the SHMR over a wide range of halo masses from  $z \sim 0.2$  to 5. At  $z \sim 0.2$ , the ratio of stellar-to-halo mass content peaks at a characteristic halo mass  $M_h = 10^{12} M_\odot$  and declines at higher and lower halo masses. This characteristic halo mass increases with redshift reaching  $M_h = 10^{12.5} M_\odot$  at  $z \sim 2.3$  and remaining flat up to  $z = 4$ . We considered the principal sources of uncertainty in our stellar mass measurements and also the variation in halo mass estimates in the literature. We show that our results are robust to these sources of uncertainty and explore likely explanation for differences between our results and those published in the literature. The steady increase in characteristic halo mass with redshift points to a scenario where cold gas inflows become progressively more important in driving star formation at high redshifts, but larger samples of massive galaxies are needed to rigorously test this hypothesis.

**Key words:** methods: statistical – galaxies: evolution – galaxies: haloes.

## 1 INTRODUCTION

Galaxy formation is a remarkably inefficient process (e.g. Silk 1977; Persic & Salucci 1992; Dayal & Ferrara 2018). This can be seen quantitatively if one compares the dark matter halo mass function (HMF) and the galaxy stellar mass function (SMF): both in low- and high-mass regimes they differ by several orders of magnitude (see e.g. Cole et al. 2001; Yang, Mo & van den Bosch 2003; Eke et al. 2006; Behroozi, Conroy & Wechsler 2010; Moster et al. 2010).

Understanding how the stellar mass content ( $M_*$ ) of a galaxy relates to the mass of its dark matter halo ( $M_h$ ) is, in fact, an alternative way of considering the problem of galaxy formation. In the local Universe, there is a ‘characteristic halo mass’ ( $M_h^{\text{peak}}$ ) at which the  $M_*/M_h$  ratio is maximized. A natural interpretation is that  $M_h^{\text{peak}}$  corresponds to the halo mass at which star formation, integrated over the entire assembly history of the galaxy, has been the most efficient (Silk, Di Cintio & Dvorkin 2013). We

consider ‘galaxy formation efficiency’ as the global process of forming stars in dark matter haloes, from the accretion of gas to the actual transformation of baryons into stars. At lower and higher halo masses, the  $M_*/M_h$  ratio decreases rapidly, presumably as a consequence of physical processes that suppress star formation in these haloes. Various mechanisms have been proposed in order to explain this inefficiency: for example, supernovae and stellar winds in low-mass haloes and active galactic nuclei (AGNs) feedback processes in more massive objects (see Silk & Mamon 2012 for a detailed review).

Although such comparisons between mass functions are phenomenological in nature (Mutch, Croton & Poole 2013) they provide useful constraints to theoretical models of galaxy formation in particular when the comparison spans a large redshift range. The advent of highly complete, mass-selected galaxy surveys (see e.g. Ilbert et al. 2013) and accurate predictions for the HMF (Tinker et al. 2008; Watson et al. 2013; Despali et al. 2016) allows us to measure the stellar-to-halo mass relationship (SHMR) of galaxies at different epochs. There are many techniques to accomplish this:

\* E-mail: louis.legrand@ias.u-psud.fr

e.g. in the ‘subhalo abundance matching’ (SHAM), the number density of galaxies (from observations) and dark matter subhaloes (from simulations) are matched to derive the SHMR at a given redshift (see e.g. Marinoni & Hudson 2002; Behroozi et al. 2010; Behroozi, Wechsler & Conroy 2013b; Behroozi et al. 2018; Moster et al. 2010; Moster, Naab & White 2013, 2018; Reddick et al. 2013). This technique can also be implemented by assuming a non-parametric monotonic relation between the luminosity or stellar mass of the observed galaxies and subhalo masses at the time of their infall on to central haloes (Conroy, Wechsler & Kravtsov 2006).

Other studies use a ‘halo occupation distribution’ modelling (HOD, see e.g. Vale & Ostriker 2004; Zheng, Coil & Zehavi 2007; Leauthaud et al. 2011; Coupon et al. 2015) where a prescription for how galaxies populate dark matter haloes can be used to simultaneously predict the number density of galaxies and their spatial distribution. In this case, lensing combined with clustering measurements can provide additional constraints on the SHMR.

However, until now, investigations of the SHMR over a large redshift range have mostly relied on heterogeneous data sets each with their own selection functions. Interpreting these results can be challenging since different biases from each survey may introduce artificial trends. In this work, we measure the SHMR and  $M_h^{\text{peak}}$  in 10 bins of redshifts between  $z = 0.2$  and  $5.5$  in a homogeneous and consistent way using the SHAM technique applied to a single data set: the COSMOS2015 galaxy catalogue (Laigle et al. 2016).

COSMOS (Scoville et al. 2007) is a  $2 \text{ deg}^2$  field with deep ultraviolet-to-infrared (UV-to-IR) coverage (see Laigle et al. 2016, and references therein). The wealth of spectroscopic observations (Lilly et al. 2007; Le Fèvre et al. 2005, 2015; Hasinger et al. 2018) means photometric redshifts can be validated even in the traditionally poorly sampled  $1 < z < 2$  redshift range (see fig. 11 of Laigle et al. and fig. 4 of Davidzon et al.). The large area of COSMOS makes it ideal to collect robust statistics of distant, massive galaxies. Moreover, exquisite IR photometry means precise stellar mass estimates can be made over a large redshift range (see e.g. Steinhardt et al. 2014; Davidzon et al. 2017). Extensive tests have been made to validate the mass completeness and the photometric redshift accuracy in COSMOS (Laigle et al. 2016; Davidzon et al. 2017). Far-IR, radio, and X-ray observations are also available to assess the crucial role of AGN (Delvecchio et al. 2017), and the quenching of distant and massive galaxies (Gozali et al. 2018).

Previously in COSMOS Leauthaud et al. (2012) used a combination of parametric abundance matching, galaxy clustering, and galaxy–galaxy lensing to derive the SHMR to  $z \sim 1$ ; galaxy–galaxy lensing measurements with COSMOS ACS data are not feasible above  $z \sim 1$ . More recently, Cowley et al. (2018) made a halo modelling analysis to derive the SHMR in the UltraVISTA ‘deep stripes’ region.

The organization of the paper is as follows. In Section 2, we introduce the observed SMF of COSMOS galaxies and discuss the principal uncertainties; we then present the Despali et al. (2016) dark matter HMF we use and our fit using a dark matter simulation to derive the HMF for the maximum mass in the history of the haloes. We also present comparisons with other mass functions for consistency checks. In Section 3, we describe our abundance matching technique, its assumptions and principal sources of uncertainties, along with our Monte Carlo Markov Chain (MCMC) fitting procedure. In Section 4, we present our results, i.e. the SHMR and its redshift evolution up to  $z \sim 5$ . We discuss the physical mechanisms that may explain our observations in Section 5.

Throughout this paper, we use the *Planck* 2015 cosmology (Planck Collaboration et al. 2016) with  $\Omega_{m,0} = 0.307$ ,  $\Omega_{\Lambda,0} = 0.691$ ,  $\Omega_{b,0} = 0.0486$ ,  $N_{\text{eff}} = 3.05$ ,  $n_s = 0.9667$ ,  $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1}) = 0.6774$ , except if noted otherwise. Stellar mass scales as  $1/h^2$  whereas halo mass scales as  $1/h$ . The notation  $\phi$  will denote a mass function. The notation  $\ln()$  refers to the natural logarithm and  $\log()$  refers to the base 10 logarithm.

## 2 MASS FUNCTIONS AND THEIR UNCERTAINTIES

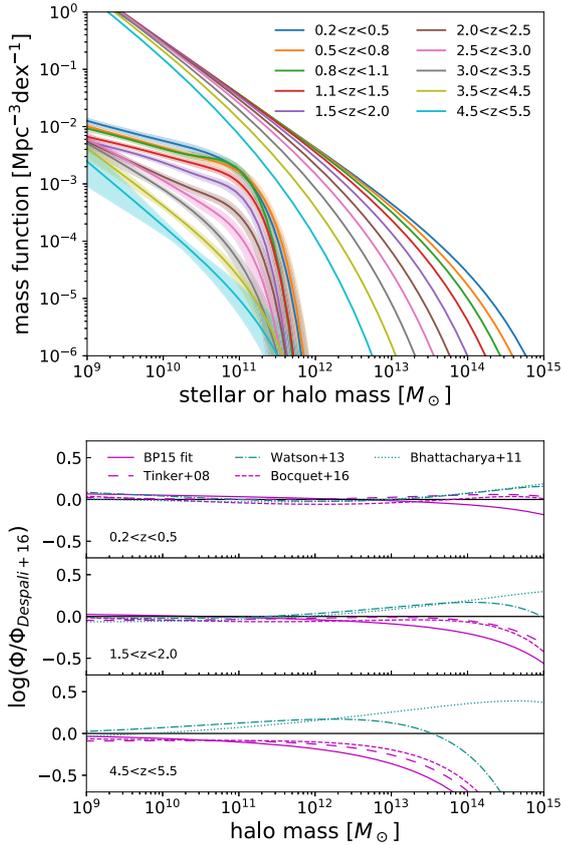
### 2.1 Stellar mass functions

The galaxy SMF corresponds to the number density per unit comoving volume of galaxies in bins of  $M_*$ . It is one of the key demographics to understand quantitatively the galaxy formation process as it describes how stellar mass is distributed in galaxies. Traditionally, the SMF has been modelled by a Schechter (1976) function, although for certain galaxy populations a combination of more than one such function may provide a better fit to observations (Binggeli, Sandage & Tammann 1988; Kelvin et al. 2014). Here, we use SMFs derived by Davidzon et al. (2017, hereafter D17) for galaxies in the UltraVISTA-Ultra deep region of the COSMOS field (see McCracken et al. 2012). The sample was constructed using the photometric catalogue of Laigle et al. (2016) which contains more than half a million galaxies with photometric redshifts ( $z_{\text{phot}}$ ) between  $z = 0.2$  and  $6$  (178 567 of them in the Ultra deep region). By restricting the analysis to the high-sensitivity region ( $K_s < 24.7 \text{ mag}$  at  $3\sigma$ ,  $\sim 0.7 \text{ mag}$  deeper than the rest of COSMOS) the effective area turns out to be  $\sim 0.5 \text{ deg}^2$ . None the less, this represents a three times larger volume than the one probed by other deep extragalactic surveys like the Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey (Grogin et al. 2011; Koekemoer et al. 2011).

Both  $z_{\text{phot}}$  and  $M_*$  are derived by fitting the galaxy spectral energy distribution (SED) with synthetic templates (see D17 for further details). The unique combination of deep optical (*Subaru*), near-IR (VISTA), and mid-IR (*Spitzer*/IRAC) observations results in a galaxy sample that is  $>90$  per cent complete at  $M_* > 10^{10} M_{\odot}$  up to  $z = 4$ ; for galaxies at  $4 < z < 6$  above that threshold, the catalogue is  $>70$  per cent complete. More generally, D17 defined a minimal mass ( $M_{*,\text{min}}$ ) as the 75 per cent completeness limit, with a redshift evolution described as  $M_{*,\text{min}}(z) = 6.3 \times 10^7 (1+z)^{2.7} M_{\odot}$ . This minimal mass is used as the lower boundary for the SMF.

D17 estimated the SMF in 10 redshift bins from  $z = 0.2$  to  $5.5$  (see Fig. 1) using three independent methods: the  $1/V_{\text{max}}$  technique (Schmidt 1968), the stepwise maximum likelihood (Efstathiou, Ellis & Peterson 1988) and the maximum likelihood method of Sandage, Tammann & Yahil (1979). These three estimators provide consistent SMF estimates. However, they are all affected by observational uncertainties ( $M_*$  and  $z_{\text{phot}}$  errors) that scatter galaxies from their original mass bin. This systematic effect, known as Eddington (1913) bias, dominates at high masses ( $M_* \gtrsim 10^{11} M_{\odot}$ ) because here galaxy number density declines exponentially; this produces an asymmetric scatter and consequently modifies the SMF profile. Depending on the ‘skewness’ and the magnitude dependence of observational errors the Eddington bias may have a strong impact also at lower masses (Grazian et al. 2015).

When fitting a Schechter (1976) function to their  $1/V_{\text{max}}$  determinations, D17 account for the Eddington bias using the method introduced in Ilbert et al. (2013). Therefore, in our work, we use the Schechter fits of D17 which should be closer to the intrinsic SMF compared to the other estimators. For consistency, we rescale these



**Figure 1.** Upper panel: our adopted stellar and HMFs. For the SMF at a given redshift (see the legend), a solid line shows our HMF (Despali et al. 2016, fitted on the Bolshoi–Planck simulation) whilst the solid line and shaded area is the SMF with the associated  $1\sigma$  uncertainty (corresponding to the best fit to  $1/V_{\max}$  points corrected for Eddington bias, Davidzon et al. 2017). Lower panel: for three sample redshift bins, the relative difference as a function of halo mass between the original (Despali et al. 2016) using the virial overdensity criterion, our Bolshoi–Planck fit (solid magenta line), and a selection of HMFs from the literature. Magenta lines show numerical simulations in which haloes are defined according to a spherical overdensity threshold (solid line: Bolshoi–Planck, long-dashed line: Tinker et al. 2008, and short-dashed line: Bocquet et al. 2016). Cyan lines show works that use an FoF algorithm (dotted line: Bhattacharya et al. 2011, dotted–dashed: Watson et al. 2013).

estimates to Planck Collaboration et al. (2016, P16) cosmology. The fitting function assumed by D17 is a double Schechter (see equation 4 in D17) at  $z < 3$  and a single Schechter function (their equation 3) above that redshift. At low redshifts, two SMF components are clearly visible (e.g. Ilbert et al. 2010), above  $z > 3$ , there is no evidence of this double Schechter profile (Wright, Driver & Robotham 2018).

The SMF error bars include both systematic and random errors including Poisson noise, cosmic variance (computed using an updated version of the software described in Moster et al. 2011) and the scatter due to errors in the SED fitting. The SMF uncertainties due to SED fitting are derived through Monte Carlo re-extraction of  $z_{\text{phot}}$  and  $M_*$  estimates according to the likelihood function of each galaxy. This procedure may be biased if the likelihood were under- or overestimated by the SED fitting code (see Dahlen et al. 2013). However, recent work with simulated photometry suggests that this should not be the case for the code used in D17 (Laigle et al., in preparation).

## 2.2 Halo mass functions

Our main reference for the dark matter HMF<sup>1</sup> is the work of Despali et al. (2016, see fig. 1). They measure the HMF using six  $N$ -body cosmological simulations with different volumes and resolutions: all of them have  $1024^3$  dark matter particles with masses ranging from  $1.94 \times 10^7$  to  $6.35 \times 10^{11} h^{-1} M_{\odot}$  and a corresponding box size from  $62.5$  to  $2000 h^{-1}$  Mpc. Haloes are identified through the ‘spherical overdensity’ algorithm (Press & Schechter 1974), i.e. each halo is a sphere with a matter density equal to the virial overdensity (see Eke, Cole & Frenk 1996) at the given redshift (which is equal to the median  $z$  of the observed SMF, see Table B2). The halo mass is defined as the sum of dark matter particles included in such a sphere.

It has been shown (see e.g. Reddick et al. 2013) that for abundance matching applications the stellar mass of galaxies is better correlated to the maximal mass the dark matter haloes have over their history ( $M_{\text{h,max}}$ ) rather than the actual mass at a given redshift. This is particularly true for subhaloes which can lose mass due to gravitational stripping by the neighbouring main halo whilst the galaxy inside will keep the same stellar mass. Reddick et al. (2013) have demonstrated that using this  $M_{\text{h,max}}$  better fits to observations such as galaxy clustering for abundance matching.

Our HMF for the maximal mass  $M_{\text{h,max}}$  are calculated using the Bolshoi–Planck simulation (Rodríguez-Puebla et al. 2016; Behroozi et al. 2018). This dark-matter-only simulation has a comoving volume of  $250 h^{-1}$  Mpc on a side with  $2048^3$  particles with a mass resolution of  $1.6 \times 10^8 h^{-1} M_{\odot}$  and uses Planck Collaboration et al. (2016) cosmology. Haloes are identified with the ROCKSTAR halo finder and masses are computed using the virial overdensity criterion of Bryan & Norman (1998). Behroozi et al. (2018) provide halo number densities for several halo mass bins and for 178 snapshots from  $z = 16$  to  $0$  for this simulation. We fit the HMF of Despali et al. (2016) using a modified version of the COLOSSUS code for these data points in the range  $0 < z < 5$  and  $10^{11} h^{-1} M_{\odot} < M_{\text{h,max}} < 10^{15} h^{-1} M_{\odot}$ . The parameters of equation 7 of Despali et al. (2016) we find are:  $A = 0.331$ ,  $a = 0.831$ , and  $p = 0.351$ . Fig. A1 shows the resulting HMF for several redshifts from  $0$  to  $5$ .

Besides the variations due to different cosmological parameters (Angulo & White 2010), it is difficult to model the uncertainties affecting the HMF. Despali et al. (2016) thoroughly discuss the implications of different density thresholds in the spherical overdensity algorithm, e.g. replacing the virial overdensity with 200 times critical ( $\rho_c$ ) or mean background ( $\rho_b$ ) density. They conclude that the virial definition leads to a ‘universal’ HMF fit, while in the case of  $\rho_c$  or  $\rho_b$  the results are more redshift dependent. A higher density threshold – e.g.  $500\rho_c$ , as often used in the literature – alters the HMF profile by decreasing the number density of the most massive systems, as some of them are now identified as a complex of smaller, independent haloes. The assumption of sphericity in the finder algorithm is less problematic since its impact on the HMF is mass-independent: accounting for haloes’ tri-axiality has only a mild effect on the HMF (Despali, Giocoli & Tormen 2014).

Other studies have investigated the impact of different halo finding algorithms which produce changes in the HMF of the order of  $\sim 10$  per cent (Knebe et al. 2011). Another potential issue is the impact of baryons (not implemented in Despali et al.) on the

<sup>1</sup>HMFs were computed using the COLOSSUS python module (Diemer 2018).

growth of dark matter haloes: Bocquet et al. (2016) show that in hydrodynamical simulations the halo number density decreases by  $\sim 15$  per cent at  $z \lesssim 0.5$  with respect to dark matter only, whereas at higher redshift, the impact of baryons is negligible.

In our analysis, we use Despali et al. mass function fitted on the Bolshoi–Planck simulation where haloes are identified using the virial overdensity criterion and where their mass is the maximal mass in their history  $M_{h, \max}$ . We also use the original version of Despali et al. (2016) HMF with halo mass defined with the virial overdensity criterion. To quantify how such a choice affects our results we consider different HMF estimates. These alternate versions are divided into two categories according to how haloes are identified. HMF estimates in the first category (Tinker et al. 2008; Bocquet et al. 2016) use the spherical overdensity definition, with halo masses defined with the  $>200\rho_b$  criterion, while the others (Bhattacharya et al. 2011; Watson et al. 2013) rely on the so-called friends-of-friends (FoF) algorithm (Davis et al. 1985). The Bolshoi–Planck fit of Despali et al. HMF is shown in the upper panel of Fig. 1 while the lower panel shows how this fit and the other HMF differ from Despali et al. in three redshift bins. At low redshifts, we find that differences are negligible, in agreement with the literature. However for  $> 10^{13} M_\odot$  haloes at  $z > 2$ , i.e. in a range barely investigated in previous work, there are 0.2–0.5 dex offsets between Despali et al. (2016) and other HMF estimates. Such a difference may be fully explained by Poisson scatter since such massive haloes are rare in the volume of cosmological simulations. We do not attempt to find the physical reasons of such a discrepancy and here we simply take the ‘inter-publication’ bias as a measure of generic HMF uncertainties (see Section 4.4).

### 3 THE STELLAR-TO-HALO MASS RELATIONSHIP

#### 3.1 The subhalo abundance matching technique

In the SHAM technique, a ‘marker’ quantity is assigned to dark matter haloes and galaxies (e.g. halo mass and stellar mass, respectively). Both haloes and galaxies are ranked according to their marker quantity, and then the latter are associated to the former by assuming a monotonic one-to-one relationship (Vale & Ostriker 2004; Conroy et al. 2006; Behroozi et al. 2010; Moster et al. 2010; Reddick et al. 2013). Here, the markers we use are the dark matter halo mass and the galaxy stellar mass.

In the hierarchical clustering scenario, small haloes accrete on to more massive ones and become ‘subhaloes’. Galaxies are classified as either ‘satellite’ (those hosted by subhaloes) or ‘central’ (those in the main halo). Since in the COSMOS2015 catalogue, there is no distinction between satellite and central galaxies to correctly perform the abundance matching we must consider all the haloes (i.e. main haloes and subhaloes) as a whole sample. For sake of simplicity, we will refer to any (sub-)halo hosting a galaxy as a ‘halo’. We do not take into account possible ‘orphan’ galaxies (i.e. satellites with no subhalo, e.g. Moster et al. 2013).

These orphan galaxies may appear when matching a catalogue of dark matter haloes with galaxies from observations (this is done in e.g. Moster et al. 2018; Behroozi et al. 2018). If the resolution of the simulation (or of the halo finder) is not precise enough, the catalogue may miss the smallest haloes and some galaxies will be unassociated. In our work, we do not use directly halo catalogues from a simulation but instead fits of a functional form of the HMF

performed on outputs of simulations. Despali et al. (2016) made sure that the fit is performed on a range of halo masses not affected by the limits of the simulation and the HMF is extrapolated below this limit to smaller masses. Campbell et al. (2018) investigated the importance of the orphan galaxies in the Bolshoi–Planck simulation. They concluded that less than 1 per cent of galaxies with  $M_* > 10^{9.5} h^{-2} M_\odot$  are orphans. As such we consider that the different HMF we use are not impacted by the resolution limits of the simulations and that the impact of orphan galaxies on the SHMR is negligible in the range of mass we consider.

The SHAM method also does not consider either the gas mass or the intracluster medium. Our sources of uncertainties are discussed in more detail in Section 3.3 (see also Behroozi et al. 2010; Campbell et al. 2018).

We perform a ‘parametric’ SHAM, assuming a functional form for the relation between  $M_*$  and  $M_h$ . Following the same formalism as in Behroozi et al. (2010), such a parametric SHMR is described by the following equation:

$$\log(M_h) = \log(M_1) + \beta \log(M_*/M_{*,0}) + \frac{(M_*/M_{*,0})^\delta}{1 + (M_*/M_{*,0})^{-\gamma}} - \frac{1}{2}. \quad (1)$$

This model has five free parameters  $M_1$ ,  $M_{*,0}$ ,  $\beta$ ,  $\delta$ , and  $\gamma$ , which determine the amplitude, the shape, and the knee of the SHMR (see Behroozi et al. 2010, for a more detailed description of the role of each parameter in shaping the SHMR). Roughly speaking, parameter values in equation (1) are adjusted during an iterative process until the HMF, converted into stellar mass through the SHMR, is in agreement with the observed SMF (see Section 3.2).

More specifically, the galaxy cumulative number density ( $N_*$ ) and the halo cumulative number density ( $N_h$ ) above a certain mass are respectively given by  $N_*(M_*) = \int_{M_*}^{+\infty} \phi_*(M) dM$  and  $N_h(M_h) = \int_{M_h}^{+\infty} \phi_h(M) dM$ , with  $\phi_*$  and  $\phi_h$  being the stellar and HMFs. The main assumptions of SHAM is that there is only one galaxy per dark matter halo and that the relation between stellar and halo masses is monotonic. As a consequence, the  $M_*$  value associated to a given  $M_h$  is the one for which  $N_*(M_*) = N_h(M_h)$ . The derivative of this equation gives the relationship between SMF, HMF, and SHMR:

$$\phi_{*, \text{conv}}(M_*) = \frac{dM_h}{dM_*} \phi_h(M_h), \quad (2)$$

where the differential term on the right-hand side can be derived from equation (1). We use the notation  $\phi_{*, \text{conv}}$  because we convolve this SMF with a lognormal distribution to account for scatter in stellar mass at fixed halo mass. The standard deviation ( $\xi$ ) of the lognormal distribution is kept as an additional free parameter; we assume that  $\xi$  is independent of the halo mass (More et al. 2009; Moster et al. 2010) but can vary with redshift. We note here that new hydrodynamical simulations like Eagle (Schaye et al. 2015) have shown that this scatter decreases from 0.25 dex at  $M_h = 10^{11} M_\odot$  to 0.12 dex at  $M_h = 10^{13} M_\odot$  (see Matthee et al. 2017). This evolution of the scatter is in agreement with latest abundance matching models (Coupon et al. 2015; Behroozi et al. 2018; Moster et al. 2018). See also fig. 9 of Gozaliasl et al. (2018). However in our analysis, we restrict ourselves to a mass-invariant scatter for simplicity.

The model SMF defined in equation (2) is then fitted to the observed one (i.e.  $\phi_{*, \text{obs}}$ ) through the procedure described in the next section.

### 3.2 Fitting procedure

To fit the model SMF to real data, a negative log-likelihood is defined as:

$$\chi^2 = \sum_i \left( \frac{\phi_{*,\text{conv}}(M_{*,i}) - \phi_{*,\text{obs}}(M_{*,i})}{\sigma_{\text{obs}}(M_{*,i})} \right)^2, \quad (3)$$

where  $\sigma_{\text{obs}}$  is the uncertainty of the observed SMF in a given stellar mass bin  $M_{*,i}$  (with the first bin starting at  $M_{*,\text{min}}$ ).

For each of the 10 redshift bins, we minimize equation (3) using an MCMC algorithm.<sup>2</sup> This algorithm allows the sampling of the parameter space in order to derive the posterior distribution for the six free parameters. We use flat conservative priors on the parameters together with 250 walkers each with a different starting point randomly selected in a Gaussian distribution around the original starting point. Our convergence criterion is based on the autocorrelation length, which is an estimate of the number of steps between which two positions of the walkers are considered uncorrelated (Goodman & Weare 2010). Our MCMC stops when the autocorrelation length has changed by less than 1 per cent and when the length of the chain is at least 50 times the autocorrelation length. As an example, our chains in the case of the HMF fitted on Bolshoi–Planck have a length between 5000 at low redshift and 25 000 in the highest redshift bin. With our 250 walkers this gives between  $1.25 \times 10^6$  and  $6.25 \times 10^6$  samples. The first steps up to two times the autocorrelation length are discarded as a burn-in phase. To speed up the computation of the posteriors, we keep only the iterations separated by a thin length which is half of the autocorrelation length.

We show in Table B2 the best fit and the 68 per cent confidence interval for the six free parameters in each of the 10 redshift bins, along with the marginalized posterior distributions in Figs B1–B3. These figures show that the parameters  $M_1$  and  $M_{*,0}$  are highly correlated. This is expected because as  $M_1$  increases,  $M_{*,0}$  should also increase.  $M_1$  and  $\beta$  are also highly correlated which may be explained by the fact that  $\log(M_*/M_{*,0})$  is negative for a large range of stellar masses so an increase of  $\beta$  is compensated by an increase of  $M_1$ . As we can see, the value of  $\delta$  is not well constrained at high redshift, because this parameter controls the high-mass slope which is not well constrained in our data.

### 3.3 Principal sources of SHAM uncertainties

There are several sources of uncertainties in the SHAM technique. A subhalo may be stripped after infall, leaving the hosted galaxy embedded in the larger, central halo. This may break the one-to-one correspondence between galaxies and dark matter haloes which is the main assumption of our method. The HOD model is a viable solution to take this into account although it would introduce an additional number of assumptions and free parameters. Moreover, observed galaxy clustering is required to constrain the HOD model parameters (e.g. Coupon et al. 2015) but such measurements are challenging at  $z > 2$  (Durkalec et al. 2015). At lower redshift ( $z \lesssim 1$ ), Leauthaud et al. (2012, see their fig. 13) have shown that  $M_h^{\text{peak}}$  measurements are consistent between HOD and SHAM measurements.

Another source of uncertainty comes from random and systematic errors in  $z_{\text{phot}}$  and  $M_*$  estimates, with the former propagating into

the  $M_*$  error in a way difficult to model (see discussion in D17). In D17, the logarithmic stellar mass uncertainty is described by a Gaussian with standard deviation  $\sigma_{M_*} = 0.35$  dex multiplied by a Lorentzian function with a parameter  $\tau$  increasing with redshift to enhance the tails of the distribution (see equation 1 of D17). These observational uncertainties which cause the Eddington bias have been corrected for in the SMF estimates we adopt (Section 2) but some caveats remain (see D17; Grazian et al. 2015). Moreover, in our fitting procedure, we consider that different stellar mass bins are uncorrelated (equation 3). This assumption is a consequence of the fact that in D17 (as the vast majority of the literature) covariance matrices are not provided for their SMF estimates. Once corrected for  $M_*$  observational uncertainties, the main source of correlations between mass bins comes from the *intrinsic* covariance between them. To avoid oversampling, we adopt a mass bin size of 0.3 dex which is comparable to the scatter in D17. We verified that this choice does not introduce any significant bias: modifying the bin size and centre (by  $\pm 0.1$  dex) results remain consistent within  $1\sigma$ .

Besides their impact on  $M_*$  estimates,  $z_{\text{phot}}$  uncertainties affect the observed SMF by scattering galaxies in the wrong redshift bin. Our binning is large enough to mitigate this given that typical  $z_{\text{phot}}$  dispersion in COSMOS2015 estimated from a large spectroscopic galaxy sample reaches  $\sigma_z \simeq 0.03(1+z)$  at  $2.5 < z < 6$ .

None the less, catastrophic  $z_{\text{phot}}$  errors in the SED fitting (e.g. due to degenerate low- and high- $z$  solution) may still be a concern. The fraction of catastrophic redshift outliers in COSMOS2015 is about 0.5 per cent at  $z < 3$  and 12 per cent at higher redshifts, so it should not introduce a significant covariance between  $z$  bins. This seems to be confirmed by test with hydrodynamical simulations (Laigle et al., in preparation).

Despite this, the impact of SED fitting systematics is still an open question which will only be resolved with next-generation surveys (e.g. large and unbiased spectroscopic samples with the Prime Focus Spectrograph at Subaru, or the James Webb Space Telescope).

However, in this work, we independently constrain parameters of equation (1) at each redshift bin without assuming a functional form for their redshift evolution (contrary to Behroozi et al. 2010). Such a redshift-independent fit reduces the overall number of assumptions in the SHMR modelling.

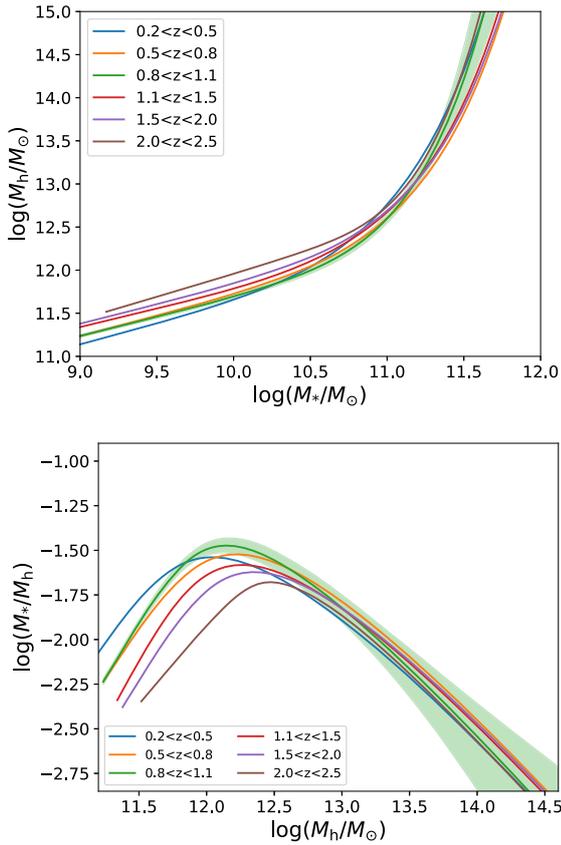
A final source of uncertainty comes from the  $M_*$  scatter we add to the galaxy-to-halo monotonic relation. This is modelled with a lognormal distribution characterized by the parameter  $\xi$  which is free to vary in the MCMC fit. This parameter is usually fixed between 0.15 and 0.20 dex (see e.g. More et al. 2009; Moster et al. 2010; Reddick et al. 2013) but in the large redshift range probed here we expect a non-negligible variation due to the evolution of galaxies' physical properties as well as observational effects. We note however that the resulting values (see Table B2) are compatible with the fixed ones assumed by the studies mentioned previously.

## 4 RESULTS

### 4.1 The stellar-to-halo mass relationship

Figs 2 and 3 show our derived SHMR fits (upper panels) and the corresponding ratio between stellar mass and halo mass (lower panels) for samples in  $0.2 < z < 2.5$  and  $2.5 < z < 5.5$  redshift intervals respectively. The SHMR and the corresponding  $1\sigma$  uncertainty are computed respectively as the 50th, the 16th, and the 84th

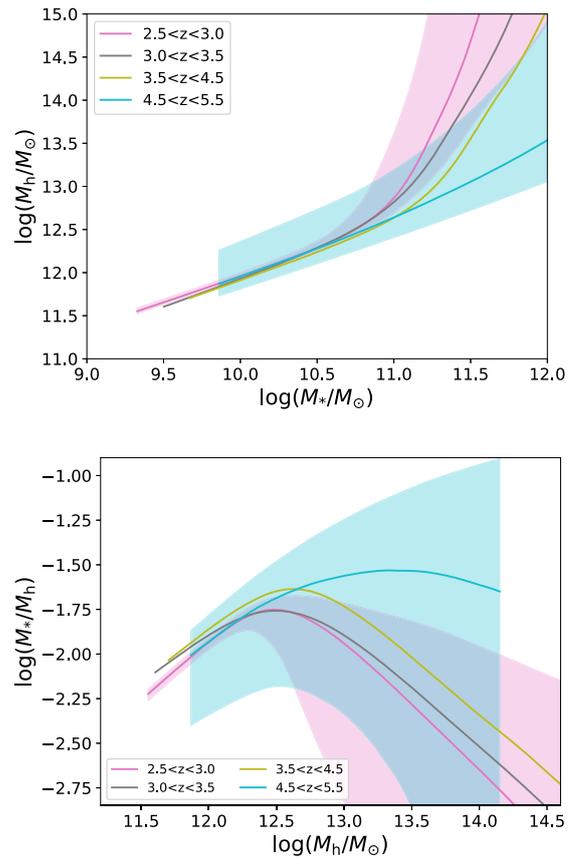
<sup>2</sup>We use the EMCEE PYTHON package (Foreman-Mackey et al. 2013).



**Figure 2.** Upper panel: SHMR from  $z = 0.2$  to  $2.5$ . Thick lines show the 50th percentile of the  $M_h$  distribution at fixed  $M_*$  computed from our MCMC runs. The coloured bands show the 16th and 84th percentile. The band is shown for the  $0.8 < z < 1.1$  redshift bin only for clarity (for other redshift bins, the uncertainty is of the same order). The limits in stellar mass for each redshift is derived from observations. Lower panel:  $M_*/M_h$  ratio derived from this SHMR.

percentile of the distribution of  $M_h$  at a given  $M_*$  in the remaining MCMC chains (Section 3.2). These uncertainties are shown as the shaded regions. Considering the stellar mass completeness of our data set (Section 2) we limit our samples to  $M_* > M_{*,\min}(z)$  and also restrict ourselves at  $M_h < 10^{15} M_\odot$  since the number density of haloes of such a mass is negligible across the whole redshift range ( $< 10^{-6} \text{Mpc}^{-3}$ ; see Mo & White 2002).

We note that the stellar mass evolution of haloes between redshift bins might sometime appear at odds with the expected stellar mass assembly. A halo with  $M_h = 10^{13} M_\odot$  has a stellar mass of  $10^{11.27} M_\odot$  at  $z = 4$ . This halo is expected to grow to a mass of  $M_h = 10^{13.5} M_\odot$  at  $z = 2.5$  where our model says the galaxy should have a stellar mass of  $10^{11.12} M_\odot$ . This effect of haloes ‘loosing’ stellar mass is a consequence of the fact that in our analysis each redshift bin is treated independently and the consistency of the model across different epochs is not guaranteed. The offset of about  $-0.15$  dex in the example above probably arises from systematic uncertainties in the SMF at high redshift, from SED fitting effects (see e.g. discussion in Stefanon et al. 2015) or from cosmic variance issues (see e.g. Davidzon et al. 2017). Besides systematic errors, statistical uncertainties are already able to explain part of the issue: running our MCMC with an SMF shifted by  $-1\sigma$  (statistical error) at  $z = 4$ , and  $+1\sigma$  at  $z = 2.5$ , the stellar mass difference in the example above is only  $-0.06$  dex.



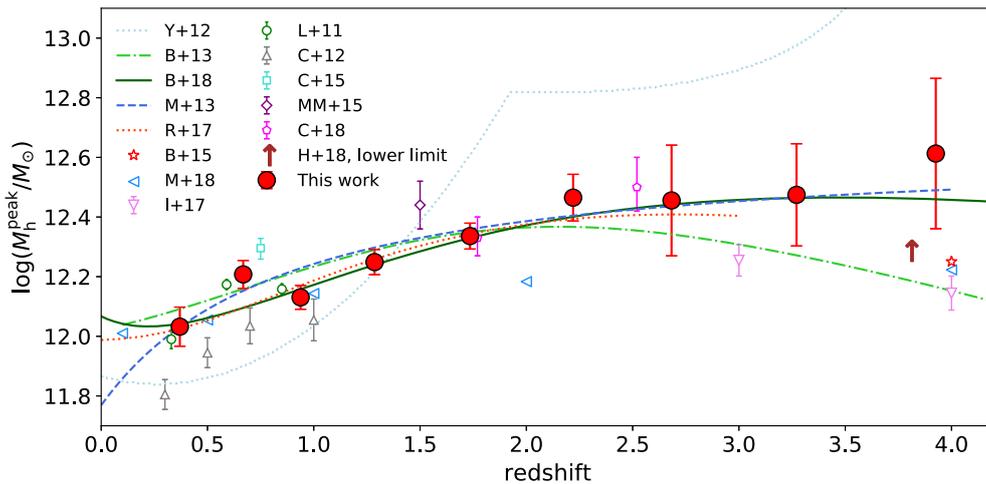
**Figure 3.** Same as Fig. 2 for redshift bins from  $z = 2.5$  to  $5.5$ . We show only uncertainties for the  $2.5 < z < 3.0$  and  $4.5 < z < 5.5$  bins for clarity.

The SHMR in the various redshift bins (upper panels of Figs 2 and 3) monotonically increases as a function of stellar mass with a changing of slope at  $\sim M_h^{\text{peak}}$ . Below the characteristic halo mass, the SHMR slope is approximately constant with redshift. Conversely, for masses above  $M_h^{\text{peak}}$ , it becomes flatter as moving towards higher redshifts (Fig. 3, upper panel) modulo the large error bars especially at  $4.5 < z < 5.5$ .

These trends are clearly illustrated also in the lower panels of Figs 2 and 3 which show  $M_*/M_h$  versus  $M_h$ . In each bin, this ratio peaks at  $M_h \simeq M_h^{\text{peak}}$  and drops by one order of magnitude at both the extremes of our halo mass range. At  $z < 0.5$ ,  $M_h^{\text{peak}} = 10^{12} M_\odot$ , with  $\log(M_*/M_h) = -1.55 \pm 0.5$ . At higher redshifts,  $M_h^{\text{peak}}$  increases steadily up to  $10^{12.5} M_\odot$  at  $z = 2$ , i.e. growing by a factor  $\sim 3$ . It then remains flat up to  $z = 4$ . At a fixed halo mass above  $M_h^{\text{peak}}$ ,  $M_*/M_h$  does not evolve, while in haloes below  $M_h^{\text{peak}}$  the ratio decreases from  $z \sim 0$  to  $2.5$ .

#### 4.2 Dependence of the peak halo mass on redshift

Fig. 4 shows the redshift evolution of the peak halo mass between  $z = 0.2$  and  $4.5$  computed from the median  $M_h^{\text{peak}}$  for all the samples retained in the MCMC (see Section 3.2). The results are reported also in Table B2. Fig. 4 also presents a compilation of recent measurements from the literature together with model predictions (lines). At  $z > 3$ , it becomes progressively more difficult to measure the position of the peak as the slopes of halo and SMFs become similar (Fig. 1). In addition at higher redshifts, there are correspondingly smaller numbers of massive galaxies in the



**Figure 4.** Peak halo mass  $M_h^{\text{peak}}$  as a function of redshift (red dots). We plot  $M_h^{\text{peak}}$  at the median redshift of each bin, rescaled to  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . All masses from other studies have been rescaled to match the  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$  cosmology. Some points from the literature have been slightly shifted along the redshift axis for clarity. We show results from Leauthaud et al. (2011, L + 11), Yang et al. (2012, Y + 12), Coupon et al. (2012, C + 12), Moster et al. (2013, M + 13), Behroozi et al. (2013b, B + 13), Behroozi & Silk (2015, B + 15), Coupon et al. (2015, C + 15), Martinez-Manso et al. (2015, MM + 15), Rodríguez-Puebla et al. (2017, R + 17), Ishikawa et al. (2017, I + 17), Cowley et al. (2018, C + 18), Harikane et al. (2018, H + 18), Moster et al. (2018, M + 18), and Behroozi et al. (2018, B + 18). The brown arrow is the lower limit for  $M_h^{\text{peak}}$  from Harikane et al. (2018, H + 18).

COSMOS volume. Nevertheless, our measurements show clearly that the peak halo mass increases steadily from  $10^{12} M_\odot$  at  $z = 0.3$  to  $10^{12.6} M_\odot$  at  $z = 4$ .

Below  $z \sim 2.5$ , there is generally a good agreement in the literature with  $M_h^{\text{peak}}$  steadily increasing as a function of redshift. We confirm this trend despite some fluctuation (e.g. at  $z \sim 0.7$ ) due to the overabundance of rich structures in COSMOS (see e.g. McCracken et al. 2015). Leauthaud et al. (2012), using a previous measurement of the COSMOS SMF at  $z < 1$ , find the same fluctuations. Leauthaud et al. perform a joint analysis of galaxy–galaxy weak lensing and galaxy clustering to fit the SHMR modelled as in Behroozi et al. (2010). Moreover, they use an HOD to describe the number of galaxies per dark matter halo, instead of assuming only a single galaxy inhabits each dark matter halo. In fact, such an assumption has only a small impact on the  $M_h^{\text{peak}}$  position given the fact that at  $\sim 10^{12} M_\odot$  most of the haloes contain only one galaxy (McCracken et al. 2015). Cowley et al. (2018) used an HOD model to derive  $M_h^{\text{peak}}$  for mass-selected sample of UltraVISTA galaxies in COSMOS at  $1.5 < z < 2$  and  $2 < z < 3$ ; their results are in good agreement with ours. Their error bars account for  $z_{\text{phot}}$  errors but not the stellar mass uncertainties; in their HMF, they apply Behroozi, Wechsler & Wu (2013a) high-redshift correction and introduce a large-scale halo bias parameter (Tinker et al. 2010).

Above  $z \gtrsim 3$ , the scatter in  $M_h^{\text{peak}}$  increases. Moster et al. (2013) and Behroozi et al. (2013b)<sup>3</sup> find different trends, i.e. an  $M_h^{\text{peak}}(z)$  function that declines (Behroozi et al. 2013b) or flattens (Moster et al. 2013) with increasing redshift. One possible explanation for the discrepancy is that Moster et al. and Behroozi et al. models are based on different observational data sets. To address this issue, Behroozi & Silk (2015) repeated Behroozi et al.’s analysis removing  $z > 5$  constraints (which in their method influence also the fit at

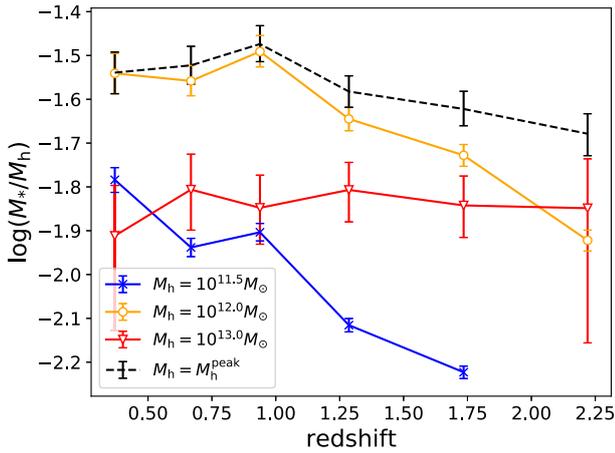
lower  $z$ ). However, this test is inconclusive as their  $M_h^{\text{peak}}$  estimate (shown as the star symbol in Fig. 4) falls between these curves.<sup>4</sup>

Our higher  $M_h^{\text{peak}}$  values with respect to Ishikawa et al. (2017) and Harikane et al. (2018) may be a consequence of our near-IR selection (a good proxy for stellar mass, see D17). Ishikawa et al. (2017) and Harikane et al. (2018) samples are selected in rest-frame UV (and a conversion to stellar mass is made through an average  $L_{\text{UV}}-M_*$  relation). Moreover their redshift classification is derived (instead of  $z_{\text{phot}}$  estimates) from a Lyman-break colour–colour selection which may result in lower levels of purity and completeness at  $z \sim 3$  (Duncan et al. 2014).

Recently, revised versions of Behroozi et al. (2013b) and of Moster et al. (2013) have been presented in Behroozi et al. (2018) and Moster et al. (2018). This new analysis differs from the former ones by following closely the evolution of individual halo–galaxy pairs through time. This results in a better understanding of the scatter of the SHMR, because this scatter results from the evolution of each halo–galaxy pairs, it is not an arbitrary scatter parameter added to the model. In Behroozi et al. (2018), the feedback model regulating star formation has significantly changed since Behroozi et al. (2013b). In the updated model, the  $M_h$  threshold at which 50 per cent of the hosting galaxies are quiescent grows from  $10^{12} M_\odot$  at  $z < 1$  up to  $\sim 10^{13} M_\odot$  at  $z = 3.5$  (see fig. 28 of Behroozi et al. 2018). As a consequence, the  $M_h^{\text{peak}}$  evolution is now in excellent agreement with both Moster et al. (2013) and our estimates. Moster et al. (2018) peak halo mass shown here corresponds to the peak in the ratio between stellar mass and baryonic mass of galaxies (the  $[M_*/M_b](M_h)$  relation). We assumed here that the ratio between baryonic mass and halo mass is a constant (equal to the universal baryon fraction), giving the same value for the peak halo mass of

<sup>3</sup>Values shown here were obtained using *Planck* cosmology instead of the published *WMAP* cosmology (Behroozi private communication). See comparison in fig. 35 of Behroozi et al. (2018).

<sup>4</sup> $M_h^{\text{peak}}(z)$  error bars are not explicitly quoted either in Behroozi et al. (2013b) or Moster et al. (2013). However, we can quantify them through the uncertainties of their SHMR models. For example in the model of Moster et al. the  $1\sigma$  confidence level of the  $M_1(z)$  parameter can be used as a proxy, leading to  $M_h^{\text{peak}}$  error bars of the same order of magnitude of ours.



**Figure 5.** Evolution of  $M_*/M_h$  as a function of redshift for fixed halo masses (solid lines) and at  $M_h \equiv M_h^{\text{peak}}$  (dashed line). Error bars are derived from Figs 2 and 3.

the  $[M_*/M_h](M_h)$  relation. The difference with our results might be explained by a dependence of the baryon fraction of haloes with mass (see Kravtsov, Nagai & Vikhlinin 2005; Davies et al. 2018).

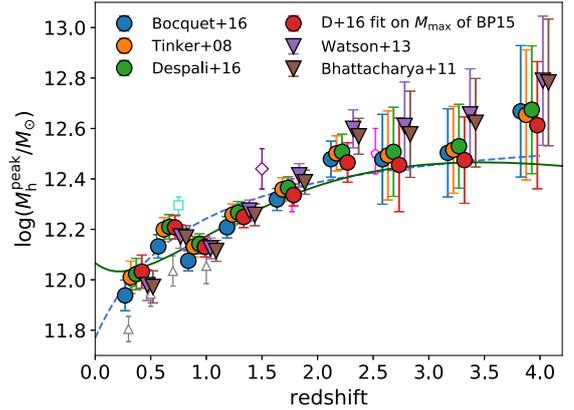
#### 4.3 Dependence of $M_*/M_h$ on redshift at fixed halo mass

Since  $M_*/M_h$  depends on host halo mass and redshift, we show in Fig. 5 this trend in more detail by computing the  $M_*/M_h$  ratio at different fixed values of halo mass. We restrict our analysis to  $z < 2.5$  because at high-mass bins ( $10^{13}M_\odot$ ) uncertainties in the  $M_*/M_h$  ratio prohibits a quantitative discussion of its evolution with redshift between  $z = 2.5$  and 5.5.

For massive haloes ( $10^{13}M_\odot$ ) the ratio is nearly constant between  $z \sim 0.2$  and 2.5, whereas at  $M_h \simeq 10^{12}M_\odot$  it increases as cosmic time goes by reaching the maximum value (about 0.03) at  $z \simeq 1$  and then remaining constant until  $z \sim 0.2$ . The redshift at which  $10^{12}M_\odot$  haloes reach the maximum  $M_*/M_h$  ratio corresponds to the epoch when  $M_h^{\text{peak}}$  is equal to their mass. Lower mass haloes, which are  $< M_h^{\text{peak}}$  across the whole redshift range, steadily increase their  $M_*/M_h$  without any peak or plateau. For instance haloes with  $M_h \simeq 10^{11.5}M_\odot$  increase their  $M_*/M_h$  ratio by a factor  $\sim 3.2$  from  $z = 2.5$  to 0.2. For comparison, Fig. 5 also shows the increase of the  $M_*/M_h$  ratio, from  $z = 2.5$  to 0.2, for haloes in a mass bin that evolves with redshift, i.e.  $M_h = M_h^{\text{peak}}(z)$ . We discuss in Section 5 the interpretation of these evolutionary trends and the implications in terms of galaxy star formation efficiency.

#### 4.4 Impact of halo mass function uncertainties

In order to estimate quantitatively how the choice of the HMF fit impacts our results, we repeat our analysis (Section 3) using different HMFs (Fig. 6). The SMF remains D17 in all the cases. Results at  $z \lesssim 2$  are consistent, whilst at higher redshifts we clearly observe the impact of halo identification techniques.  $M_h^{\text{peak}}$  values using the HMF of either Tinker et al. (2008), Bocquet et al. (2016), or Despali et al. (2016) are grouped together, as those studies all applied a spherical overdensity criterion to define haloes. Bhattacharya et al. (2011) and Watson et al. (2013) use an FoF algorithm, and the resulting  $\log(M_h^{\text{peak}}/M_\odot)$  is systematically higher by  $\sim 0.1$  dex at  $z \geq 2$ . In our study, these differences are smaller than other sources of uncertainty, but it is clear that in future larger surveys these differences may become important.



**Figure 6.** Peak halo mass ( $M_h^{\text{peak}}$ ) computed using different HMFs.  $M_h^{\text{peak}}$  redshift evolution is independently measured six times, using different HMF fits: our Bolshoi–Planck fit of Despali et al. (2016, our main reference also shown in Fig. 4); the original Despali et al. (2016), Tinker et al. (2008), Bhattacharya et al. (2011), Watson et al. (2013), and Bocquet et al. (2016). Filled circles (triangles) indicate that the halo identification has been done with a spherical overdensity (FoF) algorithm. Each set of  $M_h^{\text{peak}}(z)$  values derived for a given HMF is shifted by 0.05 in redshift for sake of clarity. SHAM method and observed SMF are the same for all estimates. Literature measurements are shown as in Fig. 4.

## 5 DISCUSSION

### 5.1 Evolution of the SHMR observed in COSMOS

To interpret our results, it is worth first recalling how the shape of the SMF changes from  $z = 5$  to 0 (Fig. 1). The number density of intermediate-mass galaxies ( $10^{9.5}M_\odot < M_* < 10^{11}M_\odot$ ) increases more rapidly compared to lower and higher masses galaxies. This causes the ‘knee’ of the SMF at  $M_* \sim 10^{11}M_\odot$  to become progressively more pronounced. In comparison halo number density evolution is nearly independent of mass so the shape of the HMF is similar between  $z = 2$  and 6 (modulo a normalization factor, see Fig. 1). The relative evolution of these two functions causes the changes in the  $M_*/M_h$  ratio.

The redshift evolution of the SMF shape is governed by several factors. On one hand, towards higher redshifts the high-mass end becomes increasingly affected by larger observational uncertainties (especially photometric redshift catastrophic failures: Caputi et al. 2015; Grazian et al. 2015). At the same time, specific physical processes control star formation around the knee of the SMF which are different from those affecting galaxies at lower masses (Peng et al. 2010). Here, we assume that most of the observational errors have been accounted for (Section 2) and consequently the redshift evolution of  $M_h^{\text{peak}}$  we measure in COSMOS is primarily driven by physical mechanisms.

The  $M_*/M_h$  ratio is usually interpreted as the comparison between the amount of star formation and dark matter accretion integrated over a halo’s lifetime. Thus, a high  $M_*/M_h$  ratio in a given  $M_h$  bin implies that those haloes have been (on average) particularly efficient in forming stars. ‘Star formation efficiency’ is used hereafter to indicate ‘galaxy formation efficiency’, i.e. the whole process of stellar mass assembly from baryon accretion to the collapse of molecular clouds inside the galaxy. In addition to the *in situ* star formation, further stellar mass can be accreted via galaxy merging. In such a framework, the dependence of the  $M_*/M_h$  ratio on halo mass and redshift can be explained by a combination of physical phenomena. Our observational constraint on  $M_h^{\text{peak}}$  can

help to understand which mechanisms, amongst those proposed in the literature, are most responsible for regulating galaxy stellar mass assembly.

$M_h^{\text{peak}}$  can be considered as the threshold above which haloes maintain a nearly constant  $M_*/M_h$  ratio across time (Fig. 5). At a fixed halo mass below  $M_h^{\text{peak}}$ , the  $M_*/M_h$  ratio increases as comoving time goes by, indicating that stellar mass has ‘kept up’ with dark matter accretion. For a fixed halo mass above  $M_h \simeq M_h^{\text{peak}}$ , host galaxies are more likely to enter in a quiescent phase (‘quenching’ the star formation) and thereafter passively evolve. Fig. 5 clearly shows this evolution with redshift for fixed halo masses. For objects with  $M_h = 10^{12} M_\odot$ , their  $M_*/M_h$  increases until  $z \sim 1$  (i.e. when  $M_h^{\text{peak}} = 1.3 - 1.6 \times 10^{12} M_\odot$ ) after which the ratio remains constant until  $z = 0.2$ . Note that we do not track the evolution of individual haloes but instead the evolution of the  $M_*/M_h$  ratio for a given halo mass. This makes the interpretation of the evolution of individual haloes with time more difficult (haloes at high redshift are not necessarily the same as haloes of the same mass at low redshift).

## 5.2 What physical mechanisms regulate star formation in our sample?

In this work, we consider primarily the redshift evolution of  $M_h^{\text{peak}}$ . Our deep near-IR observations allow us to leverage the COSMOS2015 galaxy sample to constrain that threshold up to  $z = 4$  (Fig. 4). We find that the  $M_h^{\text{peak}}(z)$  function changes slope at  $z \sim 2$ , showing a plateau at higher redshift. This implies that the threshold for massive galaxies to enter the quenching phase depends on redshift: in the early Universe quenching mechanisms are less effective for galaxies in haloes between  $10^{12}$  and  $10^{12.5} M_\odot$ . This scenario should also take into account the contribution of major and minor mergers but in the redshift and halo mass ranges of our analysis they can be considered sub-dominant (see Davidzon et al. 2018 and references therein). Therefore, in the following we will focus on quenching models affecting the *in situ* star formation.

In their cosmological hydrodynamical simulations, Gabor & Davé (2015) implement a heuristic prescription to halt star formation in systems with a large fraction of hot gas.<sup>5</sup> The condition to trigger the quenching phase, which Gabor & Davé call ‘hot halo’ mode, happens exclusively at  $M_h > 10^{12} M_\odot$  in their simulation. This halo mass threshold is in agreement with  $M_h^{\text{peak}}$ . However, Gabor & Davé (2015) carry out their analysis at  $z < 2.5$ . At higher redshift, this prescription would be in disagreement with our results. Also Behroozi et al. (2018), considering the evolution of the quiescent galaxy fraction, emphasize that a quenching recipe with a constant temperature threshold could not explain the observational trend. As the difference between a constant and a time-evolving threshold becomes more relevant in the first  $\sim 2$  Gyr after the big bang (see fig. 28 in Behroozi et al. 2018), our results are extremely useful to discriminate between these different scenarios.

The hot halo model is agnostic regarding the sub-grid physics of the simulation: gas heating can be caused by either stable virial shocks (Birnboim & Dekel 2003) or AGN feedback (see a review in Heckman & Best 2014). With respect to the former mechanism, simulations in Dekel & Birnboim (2006) show that shock heating in massive haloes becomes inefficient at high redshift because cold streams are still able to penetrate into the system and fuel star

formation (see also Dekel et al. 2009). However, despite that this trend is in general agreement with our results there are quantitative differences in the evolutionary trend. With the fiducial parameters assumed in Dekel & Birnboim (2006) the ‘critical redshift’ at which  $\sim 10^{12} M_\odot$  haloes start to form stars more efficiently is  $z_{\text{crit}} \simeq 1.5$ . Moreover, according to their model  $M_h^{\text{peak}}$  should keep increasing at  $z > z_{\text{crit}}$  instead of plateauing.

Quenching models more compatible with our observational results have been presented e.g. in Feldmann & Mayer (2015). Under the assumption that gas inflow (thus star formation) is strongly correlated to dark matter accretion, the authors note that at  $z > 2$  massive haloes are still collapsing fast and dark matter filaments efficiently funnel cool gas into the galaxy. At  $z \lesssim 2$ , those haloes should enter in a phase of slower accretion that eventually impedes star formation by gas starvation. However, we caution that they study single galaxies in cosmological zoom-in simulations: a larger sample may show considerable variance in the redshift marking the transition between the two dark matter accretion phases. In addition, we emphasize that not only the accretion rate but also the cooling timescale is a strong function of redshift. Gas density follows the overall matter density of the Universe, evolving as  $\propto(1+z)^3$ . Since the post-shock cooling time is proportional to gas density, it would be significantly shorter at higher redshift. On the other hand, this argument in absence of more complex factors should lead to a steeper, monotonic increase of  $M_h^{\text{peak}}$  that we do not observe.

As mentioned above, AGN feedback at high redshifts can also regulate galaxy star formation and explain our observed  $M_h^{\text{peak}}$  trend. AGN activity at high redshift is expected to be almost exclusively in quasar mode (e.g. Silk & Rees 1998) with powerful outflows that can heat or even expel gas. However such radiative feedback has shown to be inefficient in hydrodynamical ‘zoom-in’ simulations at  $z \sim 6$  (e.g. Costa et al. 2014). Observations also indicate that high- $z$  quasars do not prevent significant reservoirs of cold gas from fuelling star formation (e.g. Maiolino et al. 2012; Ciccone et al. 2014). Therefore, star formation in massive haloes can proceed for 2–3 Gyr after the big bang without being significantly affected by AGN activity, in agreement with our observations. At later times, perturbations to cold filamentary accretion can starve galaxies of their gas supplies (Dubois et al. 2013).

A deeper understanding of the role played by AGN comes from studying their co-evolution with supermassive black hole (BH). Beckmann et al. (2017) show that once renormalized for the ratio between the BH mass ( $M_{\text{BH}}$ ) and the virial mass of the halo, the impact of AGN feedback is the same from  $z = 0$  to 5. According to their hydrodynamical simulations (from the HORIZON-AGN suite, Dubois et al. 2014) this process is able to suppress galaxy stellar mass assembly when  $M_{\text{BH}}/M_h > 4 \times 10^{-5}$ . In first approximation, this critical threshold is in good agreement with the one that can be derived from COSMOS if the critical BH mass ( $M_{\text{BH,crit}}$ ) is correlated with  $M_h^{\text{peak}}$ . Assuming a BH-to-stellar mass ratio of  $2 \times 10^{-3}$  (Marconi & Hunt 2003) we can write

$$\frac{M_{\text{BH,crit}}}{M_h^{\text{peak}}} = \frac{M_{\text{BH}}}{M_*} \times \left(\frac{M_*}{M_h}\right)^{\text{peak}} = 2 \times 10^{-3} \times 10^{-1.7 \pm 0.1}, \quad (4)$$

which gives  $3\text{--}5 \times 10^{-5}$  including the variation in the  $M_*/M_h$  ratio calculated at  $M_h \equiv M_h^{\text{peak}}$  (see Figs 2, 3, and 5). At least at  $z < 2$ , the antihierarchical growth of the BH mass function (Marconi et al. 2004; Shankar, Weinberg & Miralda-Escudé 2009, 2013) implies that more massive BHs form earlier, so  $M_h^{\text{peak}}$  must also increase (as we find in COSMOS) in order to keep the ratio constant. In other words, if we assume that the quenching threshold  $M_{\text{BH,crit}}/M_h^{\text{peak}}$  is universal, BH formation models can use the COSMOS SHMR as an

<sup>5</sup>Namely, their code prevents gas cooling in FoF structures by setting the circumgalactic gas equal to the virial temperature. This condition is triggered when a structure has 60 per cent of gas particles with a temperature  $> 10^{5.4}$  K (Kereš et al. 2005).

indirect constraint. Part of the redshift evolution may also be due to the ratio between  $M_{\text{BH}}$  and  $M_*$ . In equation (4), we used a constant value but other studies indicate that such a relationship varies depending on the galaxy bulge component (Reines & Volonteri 2015). However, in the HORIZON-AGN simulation this quantity has been shown to remain constant ( $\sim 2 \times 10^{-3}$ ) up to  $z = 3$  (Volonteri et al. 2016).

Recently, Glazebrook et al. (2017) reported a massive ( $M_* = 1.7 \times 10^{11} M_\odot$ ) and quiescent galaxy at a spectroscopic redshift of  $z = 3.717$ . This observation suggests a scenario where in the early Universe dark matter haloes are hosting massive star-forming galaxies and that the quenching of star formation appears as early as  $z \sim 4$ . According to our SHMR relation, this stellar mass corresponds to a halo mass of  $M_h = \sim 10^{12.5} M_\odot$  so around our value of  $M_h^{\text{peak}}$  for  $z = 4$ . This observation is in agreement with our argument that  $M_h^{\text{peak}}$  is the characteristic mass for haloes currently undergoing quenching.

Depending on their location within the cosmic web (filaments, nodes, voids) haloes with similar masses may experience different accretion histories (De Lucia et al. 2012). One key idea in this context is ‘cosmic web detachment’ (Aragon-Calvo, Neyrinck & Silk 2016): galaxies tied to nodes or filaments are removed from their original location by interaction with another galaxy. After the detachment gas supply – and then star formation – becomes less efficient. Aragon-Calvo et al. (2016) suggest that massive haloes are the first to detach, whereas less massive haloes  $0.1 - 3 \times 10^{10} h^{-1} M_\odot$  are still part of the cosmic web today. It is difficult to test this scenario beyond the local Universe because precise measurements of the SMF are required in addition to higher order statistics (e.g. three-point correlation functions). We emphasize that COSMOS is the ideal laboratory to test the impact of large-scale environment in the models mentioned above, because the cosmic structure of this field has been reconstructed at least up to  $z \sim 1$  (Darvish et al. 2014; Laigle et al. 2018). We aim to perform such an analysis in a future work.

In summary, we have described different physical processes which could explain our observed trends of  $M_h^{\text{peak}}$  and the SHMR with redshift. Of course, in the real Universe the truth is likely to be some combination of these mechanisms. But based on this discussion, the physical processes at work in results seem to be best understood as a combination of cold-flow accretion and AGN feedback combined with antihierarchical growth of the BH mass function, with the precise role of evolutionary and environmental effects yet to be determined.

## 6 CONCLUSIONS

We have used an SHAM technique combined with precise SMF measurements in COSMOS to make a new measurement of the SHMR from  $z \sim 0$  to 5. We accounted for the principal sources of uncertainties in our stellar mass measurements and photometric redshifts. We also tested the impact of HMF uncertainties on the resulting SHMR. At  $z \sim 0.2$ , we found that the ratio of mass in stars to dark matter halo mass ( $M_*/M_h$ ) peaks at a halo mass of  $10^{12.05 \pm 0.07} M_\odot$ . This peak mass increases steadily to  $10^{12.48 \pm 0.08} M_\odot$  at  $z \sim 2.3$ , and remains almost constant up to  $z = 4$ .

By comparing our results to studies that rely on models accounting for both central and satellite galaxies, we have shown that at least at  $z < 2$ , the distinction between central and satellite galaxies has only a limited impact on the peak halo mass  $M_h^{\text{peak}}$ . A complete modelling including satellite galaxies is left to a future work.

We found that the  $M_*/M_h$  ratio has little dependence on redshift for haloes more massive than  $M_h^{\text{peak}}$ , but strongly depends on redshift for less-massive haloes, consistent with the picture that the star formation has been quenched in massive haloes and continues in less massive haloes. We showed that the evolution of the shape of the SMF has a strong impact on the SHMR: the change in the position of the knee of the SMF is responsible for the shift in the value of  $M_h^{\text{peak}}$ . Accurate SMF estimations at high redshift for massive galaxies are needed to constrain the SHMR. We also show how mass function uncertainties can influence our measurements of  $M_h^{\text{peak}}$ .

We discussed qualitatively which physical processes control the SHMR and  $M_h^{\text{peak}}$ , which we interpret as the characteristic mass of quenched haloes. We speculate that this evolutions can either be related to AGN feedback or to environmental effects such as cold gas inflows at high redshift and cosmic web detachment.

Our study is based on a phenomenological model and as such can provide no direct information concerning the physical processes acting inside haloes. Next-generation hydrodynamical simulations will allow us to better understand the small-scale physical processes acting inside dark matter haloes and determine what physical effects control star formation. In the next few years, the combined 20 deg<sup>2</sup> Spitzer–Euclid legacy and Hawaii-2-0 surveys on the Euclid deep fields will provide much better constraints on the massive end of the SMF at high redshifts. Precise photometric redshifts will allow us to investigate in detail the role of environment and in particular the ‘cosmic web’ role in shaping galaxy and dark matter evolution.

For future surveys like Euclid, errors on the cosmological figure of merit will be dominated by systematic errors. For this reason, it is essential to understand the interplay between baryons and dark matter on small scales and the uncertainties present in estimates of the HMF.

## ACKNOWLEDGEMENTS

The COSMOS team in France acknowledges support from the Centre National d’Études Spatiales. LL acknowledges financial support from Euclid Consortium and CNES. HJM acknowledges financial support from the ‘Programme national cosmologie et galaxies’ (PNCG) and the DIM-ACAV. OI acknowledge funding of the French Agence Nationale de la Recherche for the ‘SAGACE’ project. Support for this work was also provided by NASA. The Cosmic Dawn Center is funded by the DNRF. This research is also partly supported by the Centre National d’Études Spatiales (CNES). The authors wish to thank P. Behroozi, J. Silk, M. Lehnert, G. Mamon, N. Malavasi, R. M. J. Janssen, and G. Despali for their useful comments on an earlier version of this manuscript. The authors would also like to thank the referee for a thorough and constructive report which improved this paper. This work is based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under ESO programme ID 179.A-2005 and on data products produced by TERAPIX and the Cambridge Astronomy Survey Unit on behalf of the UltraVISTA consortium. It is also based on observations and archival data made with the *Spitzer Space Telescope*, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA.

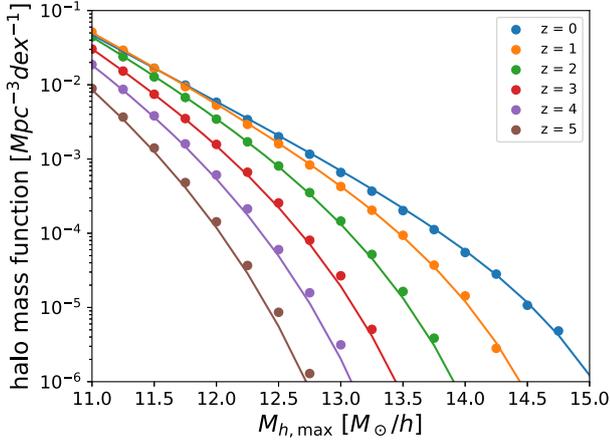
## REFERENCES

- Angulo R. E., White S. D. M., 2010, *MNRAS*, 405, 143  
 Aragon-Calvo M. A., Neyrinck M. C., Silk J., 2016, preprint (arXiv:1607.07881)  
 Beckmann R. S. et al., 2017, *MNRAS*, 472, 949

- Behroozi P., Wechsler R., Hearin A., Conroy C., 2019, MNRAS accepted
- Behroozi P. S., Silk J., 2015, *ApJ*, 799, 32
- Behroozi P. S., Conroy C., Wechsler R. H., 2010, *ApJ*, 717, 379
- Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013a, *ApJ*, 762, 109
- Behroozi P. S., Wechsler R. H., Conroy C., 2013b, *ApJ*, 770, 57
- Bhattacharya S., Heitmann K., White M., Lukić Z., Wagner C., Habib S., 2011, *ApJ*, 732, 122
- Binggeli B., Sandage A., Tammann G. A., 1988, *ARA&A*, 26, 509
- Birboim Y., Dekel A., 2003, *MNRAS*, 345, 349
- Bocquet S., Saro A., Dolag K., Mohr J. J., 2016, *MNRAS*, 456, 2361
- Bryan G. L., Norman M. L., 1998, *ApJ*, 495, 80
- Campbell D., van den Bosch F. C., Padmanabhan N., Mao Y.-Y., Zentner A. R., Lange J. U., Jiang F., Villarreal A., 2018, *MNRAS*, 477, 359
- Caputi K. I. et al., 2015, *ApJ*, 810, 73
- Cicone C. et al., 2014, *A&A*, 562, A21
- Cole S. et al., 2001, *MNRAS*, 326, 255
- Conroy C., Wechsler R. H., Kravtsov A. V., 2006, *ApJ*, 647, 201
- Costa T., Sijacki D., Trenti M., Haehnelt M., 2014, *MNRAS*, 439, 2146
- Coupon J. et al., 2012, *A&A*, 542, A5
- Coupon J. et al., 2015, *MNRAS*, 449, 1352
- Cowley W. I. et al., 2018, *ApJ*, 853, 69
- Dahlen T. et al., 2013, *ApJ*, 775, 93
- Darvish B., Sobral D., Mobasher B., Scoville N. Z., Best P., Sales L. V., Smail I., 2014, *ApJ*, 796, 51
- Davidzon I. et al., 2017, *A&A*, 605, A70 (D17)
- Davidzon I., Ilbert O., Faisst A. L., Sparre M., Capak P. L., 2018, *ApJ*, 852, 107
- Davies J. J., Crain R. A., McCarthy I. G., Oppenheimer B. D., Schaye J., Schaller M., McAlpine S., 2018, *MNRAS*, 485, 3783
- Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *ApJ*, 292, 371
- Dayal P., Ferrara A., 2018, *Phys. Rep.*, 780, 1
- De Lucia G., Weinmann S., Poggianti B. M., Aragón-Salamanca A., Zaritsky D., 2012, *MNRAS*, 423, 1277
- Dekel A. et al., 2009, *Nature*, 457, 451
- Dekel A., Birboim Y., 2006, *MNRAS*, 368, 2
- Delvecchio I. et al., 2017, *A&A*, 602, A3
- Despali G., Giocoli C., Tormen G., 2014, *MNRAS*, 443, 3208
- Despali G., Giocoli C., Angulo R. E., Tormen G., Sheth R. K., Baso G., Moscardini L., 2016, *MNRAS*, 456, 2486
- Diemer B., 2018, *ApJS*, 239, 35
- Dubois Y. et al., 2014, *MNRAS*, 444, 1453
- Dubois Y., Pichon C., Devriendt J., Silk J., Haehnelt M., Kimm T., Slyz A., 2013, *MNRAS*, 428, 2885
- Duncan K. et al., 2014, *MNRAS*, 444, 2960
- Durkalec A. et al., 2015, *A&A*, 583, A128
- Eddington A. S., 1913, *MNRAS*, 73, 359
- Efstathiou G., Ellis R. S., Peterson B. A., 1988, *MNRAS*, 232, 431
- Eke V. R., Cole S., Frenk C. S., 1996, *MNRAS*, 282
- Eke V. R., Baugh C. M., Cole S., Frenk C. S., Navarro J. F., 2006, *MNRAS*, 370, 1147
- Feldmann R., Mayer L., 2015, *MNRAS*, 446, 1939
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Gabor J. M., Davé R., 2015, *MNRAS*, 447, 374
- Glazebrook K. et al., 2017, *Nature*, 544, 71
- Goodman J., Weare J., 2010, *Commun. Appl. Math. Comput. Sci.*, 5, 65
- Gozali G. et al., 2018, *MNRAS*, 475, 2787
- Grazian A. et al., 2015, *A&A*, 575, A96
- Grogin N. A. et al., 2011, *ApJS*, 197, 35
- Harikane Y. et al., 2018, *PASJ*, 70, S11
- Hasinger G. et al., 2018, *ApJ*, 858, 77
- Heckman T. M., Best P. N., 2014, *ARA&A*, 52, 589
- Ilbert O. et al., 2010, *ApJ*, 709, 644
- Ilbert O. et al., 2013, *A&A*, 556, A55
- Ishikawa S., Kashikawa N., Toshikawa J., Tanaka M., Hamana T., Niino Y., Ichikawa K., Uchiyama H., 2017, *ApJ*, 841, 8
- Kelvin L. S. et al., 2014, *MNRAS*, 444, 1647
- Kereš D., Katz N., Weinberg D. H., Davé R., 2005, *MNRAS*, 363, 2
- Knebe A. et al., 2011, *MNRAS*, 415, 2293
- Koekemoer A. M. et al., 2011, *ApJS*, 197, 36
- Kravtsov A. V., Nagai D., Vikhlinin A. A., 2005, *ApJ*, 625, 588
- Laigle C. et al., 2016, *ApJS*, 224, 24
- Laigle C. et al., 2018, *MNRAS*, 474, 5437
- Le Fèvre O. et al., 2005, *A&A*, 439, 845
- Le Fèvre O. et al., 2015, *A&A*, 576, A79
- Leauthaud A. et al., 2012, *ApJ*, 744, 159
- Leauthaud A., Tinker J., Behroozi P. S., Busha M. T., Wechsler R. H., 2011, *ApJ*, 738, 45
- Lilly S. J. et al., 2007, *ApJS*, 172, 70
- Maiolino R. et al., 2012, *MNRAS*, 425, L66
- Marconi A., Hunt L. K., 2003, *ApJ*, 589, L21
- Marconi A., Risaliti G., Gilli R., Hunt L. K., Maiolino R., Salvati M., 2004, *MNRAS*, 351, 169
- Marinoni C., Hudson M. J., 2002, *ApJ*, 569, 101
- Martinez-Manso J., Gonzalez A. H., Ashby M. L. N., Stanford S. A., Brodwin M., Holder G. P., Stern D., 2015, *MNRAS*, 446, 169
- Matthee J., Schaye J., Crain R. A., Schaller M., Bower R., Theuns T., 2017, *MNRAS*, 465, 2381
- McCracken H. J. et al., 2012, *A&A*, 544, A156
- McCracken H. J. et al., 2015, *MNRAS*, 449, 901
- More S., van den Bosch F. C., Cacciato M., Mo H. J., Yang X., Li R., 2009, *MNRAS*, 392, 801
- Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, *ApJ*, 710, 903
- Moster B. P., Somerville R. S., Newman J. A., Rix H.-W., 2011, *ApJ*, 731, 113
- Moster B. P., Naab T., White S. D. M., 2013, *MNRAS*, 428, 3121
- Moster B. P., Naab T., White S. D. M., 2018, *MNRAS*, 477, 1822
- Mo H. J., White S. D. M., 2002, *MNRAS*, 336, 112
- Mutch S. J., Croton D. J., Poole G. B., 2013, *MNRAS*, 435, 2445
- Peng Y.-j. et al., 2010, *ApJ*, 721, 193
- Persic M., Salucci P., 1992, *MNRAS*, 258, 14P
- Planck Collaboration et al., 2016, *A&A*, 594, A13
- Press W. H., Schechter P., 1974, *ApJ*, 187, 425
- Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, *ApJ*, 771, 30
- Reines A. E., Volonteri M., 2015, *ApJ*, 813, 82
- Rodríguez-Puebla A., Behroozi P., Primack J., Klypin A., Lee C., Hellinger D., 2016, *MNRAS*, 462, 893
- Rodríguez-Puebla A., Primack J. R., Avila-Reese V., Faber S. M., 2017, *MNRAS*, 470, 651
- Sandage A., Tammann G. A., Yahil A., 1979, *ApJ*, 232, 352
- Schaye J. et al., 2015, *MNRAS*, 446, 521
- Schechter P., 1976, *ApJ*, 203, 297
- Schmidt M., 1968, *ApJ*, 151, 393
- Scoville N. et al., 2007, *ApJS*, 172, 1
- Shankar F., Weinberg D. H., Miralda-Escudé J., 2009, *ApJ*, 690, 20
- Shankar F., Weinberg D. H., Miralda-Escudé J., 2013, *MNRAS*, 428, 421
- Silk J., 1977, *ApJ*, 211, 638
- Silk J., Mamon G. A., 2012, *Res. Astron. Astrophys.*, 12, 917
- Silk J., Rees M. J., 1998, *A&A*, 331, L1
- Silk J., Di Cintio A., Dvorkin I., 2013, preprint ([arXiv:1312.0107](https://arxiv.org/abs/1312.0107))
- Stefanon M. et al., 2015, *ApJ*, 803, 11
- Steinhardt C. L. et al., 2014, *ApJ*, 791, L25
- Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, *ApJ*, 688, 709
- Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, *ApJ*, 724, 878
- Vale A., Ostriker J. P., 2004, *MNRAS*, 353, 189
- Volonteri M., Dubois Y., Pichon C., Devriendt J., 2016, *MNRAS*, 460, 2979
- Watson W. A., Iliev I. T., D'Aloisio A., Knebe A., Shapiro P. R., Yepes G., 2013, *MNRAS*, 433, 1230
- Wright A. H., Driver S. P., Robotham A. S. G., 2018, *MNRAS*, 480, 3491
- Yang X., Mo H. J., van den Bosch F. C., 2003, *MNRAS*, 339, 1057
- Yang X., Mo H. J., van den Bosch F. C., Zhang Y., Han J., 2012, *ApJ*, 752, 41
- Zheng Z., Coil A. L., Zehavi I., 2007, *ApJ*, 667, 760

**APPENDIX A: FITTING THE BOLSHOI-PLANCK HMF FOR  $M_{\text{MAX}}$** 

Fig. A1 shows the fit of the Despali et al. (2016) HMF on the halo number densities of the Bolshoi–Planck simulation, using the maximal mass in the history of the haloes. See Section 2.2.



**Figure A1.** Points show halo densities obtained from the Bolshoi–Planck simulation for different redshift snapshots. The fit of the Despali et al. (2016) HMF on this data points is shown as the plain lines.

**APPENDIX B: MCMC BEST FIT PARAMETERS**

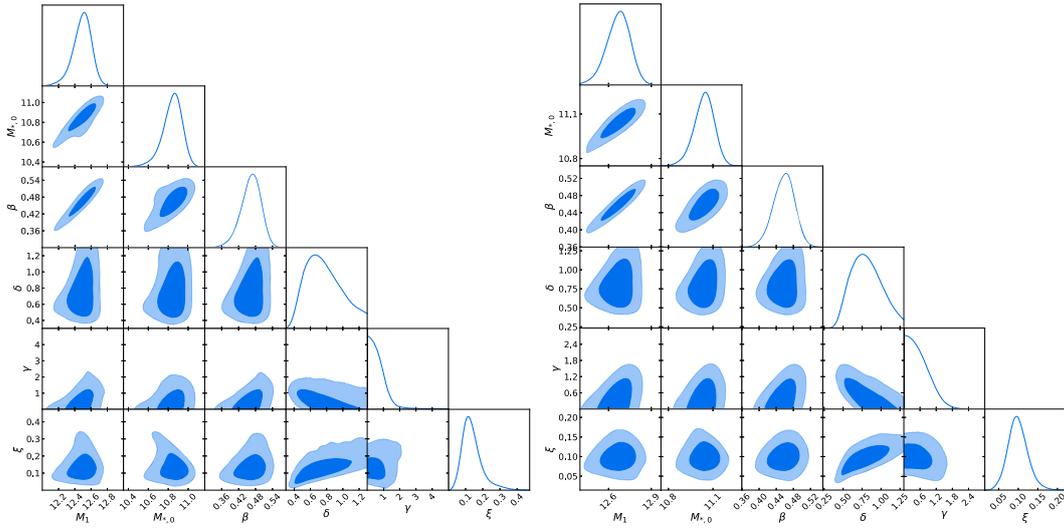
In this Appendix, we provide further details about the COSMOS2015 galaxy SMF (median  $z$  and limiting stellar mass of each bin, see Table B1) and the best-fitting parameters of equation (1) resulting from our MCMC method (see Table B2). In addition, Figs B1–B3 show the MCMC posterior distributions in the 10 redshift bins independently considered in this analysis.

**Table B1.** Median redshift of each redshift bin and limiting stellar mass of COSMOS survey as defined in D17.

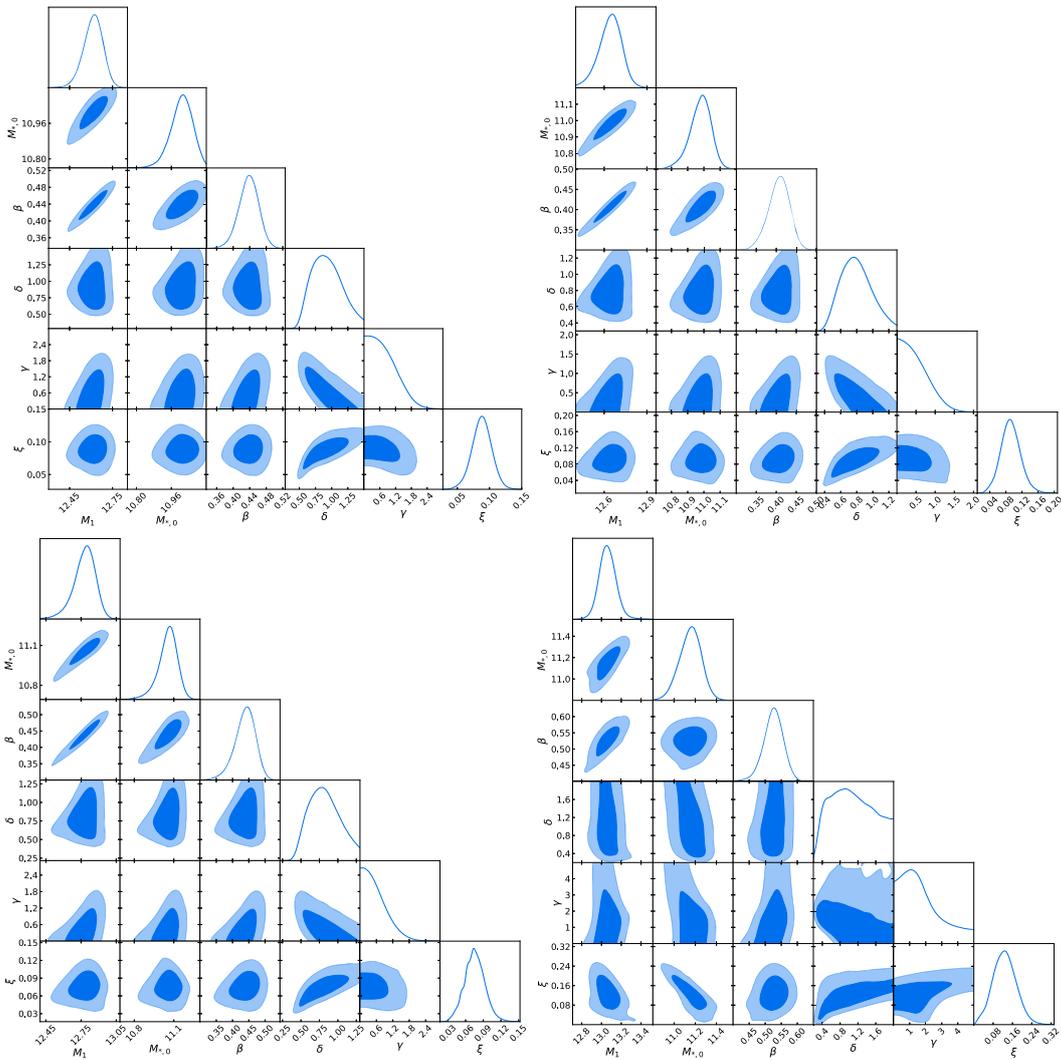
Redshift bin	Median $z$	$\log(M_{*,\text{lim}}/M_{\odot})$
(0.2, 0.5]	0.370	8.17
(0.5, 0.8]	0.668	8.40
(0.8, 1.1]	0.938	8.58
(1.1, 1.5]	1.29	8.77
(1.5, 2.0]	1.74	8.98
(2.0, 2.5]	2.22	9.17
(2.5, 3.0]	2.68	9.32
(3.0, 3.5]	3.27	9.50
(3.5, 4.5]	3.93	9.67
(4.5, 5.5]	4.80	9.86

**Table B2.** Best-fitting parameters for the 10 redshift bins with their 68 per cent confidence intervals, and  $M_{\text{h}}^{\text{peak}}$  recovered from the best-fitting SHMR with its 68 per cent confidence interval.

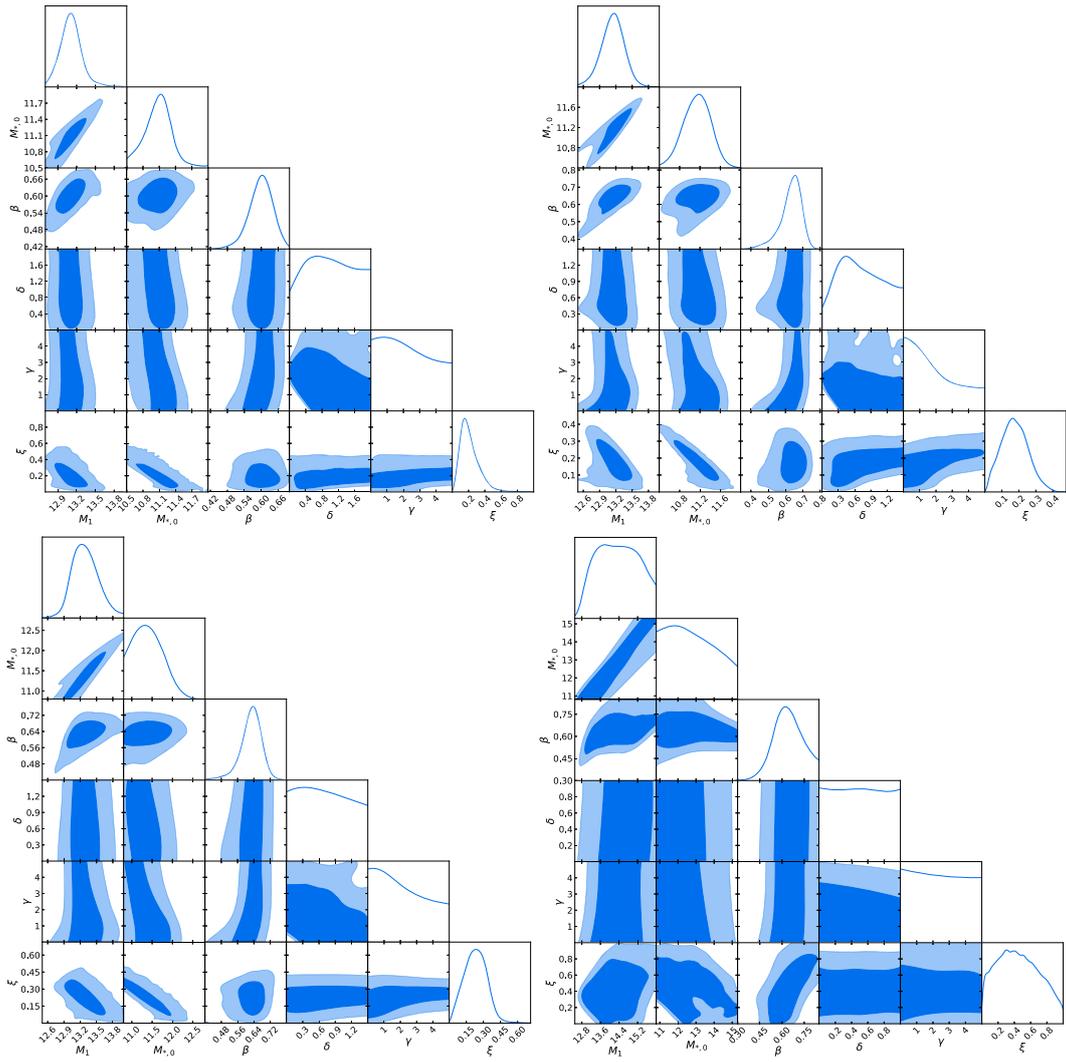
Redshift bin	$\log(M_1/M_{\odot})$	$\log(M_{*,0}/M_{\odot})$	$\beta$	$\delta$	$\gamma$	$\xi$	$\log(M_{\text{h}}^{\text{peak}}/M_{\odot})$
[0.2, 0.5]	$12.49^{+0.13}_{-0.094}$	$10.84^{+0.11}_{-0.077}$	$0.463^{+0.040}_{-0.030}$	$0.77^{+0.16}_{-0.29}$	$<0.802$	$0.138^{+0.034}_{-0.066}$	$12.05 \pm 0.07$
[0.5, 0.8]	$12.668^{+0.089}_{-0.074}$	$11.039^{+0.074}_{-0.060}$	$0.458^{+0.026}_{-0.023}$	$0.81^{+0.17}_{-0.24}$	$<0.723$	$0.099^{+0.022}_{-0.027}$	$12.22 \pm 0.05$
[0.8, 1.1]	$12.614^{+0.073}_{-0.060}$	$11.006^{+0.056}_{-0.042}$	$0.437^{+0.025}_{-0.022}$	$0.93^{+0.19}_{-0.28}$	$<0.955$	$0.088 \pm 0.015$	$12.14 \pm 0.04$
[1.1, 1.5]	$12.642^{+0.086}_{-0.069}$	$10.978^{+0.072}_{-0.054}$	$0.407^{+0.029}_{-0.023}$	$0.80^{+0.16}_{-0.23}$	$<0.629$	$0.092^{+0.023}_{-0.025}$	$12.26 \pm 0.04$
[1.5, 2.0]	$12.78^{+0.10}_{-0.072}$	$11.053^{+0.080}_{-0.055}$	$0.438^{+0.035}_{-0.026}$	$0.82^{+0.17}_{-0.25}$	$<0.724$	$0.075 \pm 0.017$	$12.35 \pm 0.04$
[2.0, 2.5]	$13.062^{+0.078}_{-0.087}$	$11.15^{+0.11}_{-0.095}$	$0.525^{+0.033}_{-0.027}$	$1.09^{+0.36}_{-0.68}$	$<2.08$	$0.128^{+0.045}_{-0.050}$	$12.48 \pm 0.08$
[2.5, 3.0]	$13.11 \pm 0.18$	$11.09 \pm 0.25$	$0.598^{+0.045}_{-0.036}$	$1.01^{+0.55}_{-0.72}$	–	$0.216^{+0.061}_{-0.14}$	$12.47 \pm 0.19$
[3.0, 3.5]	$13.14^{+0.22}_{-0.20}$	$11.14 \pm 0.27$	$0.631^{+0.071}_{-0.038}$	$0.73^{+0.35}_{-0.54}$	$<2.47$	$0.176^{+0.074}_{-0.085}$	$12.49 \pm 0.17$
[3.5, 4.5]	$13.30^{+0.20}_{-0.27}$	$11.41^{+0.28}_{-0.46}$	$0.625^{+0.056}_{-0.039}$	–	$<2.93$	$0.231 \pm 0.099$	$12.63 \pm 0.25$
[4.5, 5.5]	$14.35^{+0.89}_{-1.0}$	$<13.5$	$0.642^{+0.094}_{-0.11}$	–	–	$0.45^{+0.22}_{-0.34}$	$13.35 \pm 0.54$



**Figure B1.** 1D and 2D marginalized distributions for the six free parameters. Solid contours give the 68 and 95 per cent confidence intervals. Left-hand panel is for redshift bin [0.2, 0.5], and right-hand panel is for redshift bin [0.5, 0.8].



**Figure B2.** Same as B1 for (from left to right and top to bottom) [0.8, 1.1], [1.1, 1.5], [1.5, 2], and [2, 2.5] redshift bins.



**Figure B3.** Same as B1 for (from left to right and top to bottom) [2.5, 3], [3, 3.5], [3.5, 4.5], and [4.5, 5.5] redshift bins.

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.

# High resolution tomography for galaxy spectroscopic surveys with Angular Redshift Fluctuations

L. Legrand<sup>1</sup>, C. Hernández-Monteagudo<sup>2</sup>, M. Douspis<sup>1</sup>, N. Aghanim<sup>1</sup>, and Raúl E. Angulo<sup>3,4</sup>

<sup>1</sup> Université Paris-Saclay, CNRS, Institut d'astrophysique spatiale, 91405, Orsay, France.  
e-mail: louis.legrand@ias.u-psud.fr

<sup>2</sup> Centro de Estudios de Física del Cosmos de Aragón (CEFCA), Unidad Asociada al CSIC, Plaza San Juan, 1, planta 2, E-44001, Teruel, Spain

<sup>3</sup> Donostia International Physics Centre (DIPC), Paseo Manuel de Lardizabal 4, E-20018 Donostia-San Sebastian, Spain

<sup>4</sup> IKERBASQUE, Basque Foundation for Science, E-48013, Bilbao, Spain

July 28, 2020

## ABSTRACT

In the context of next generation spectroscopic galaxy surveys, new observables of the distribution of matter are currently being developed. Among these we investigate the angular redshift fluctuations (ARF), which probe the information contained in the projected redshift distribution of galaxies. Relying on the Fisher formalism, we show how ARF will provide complementary cosmological information compared to traditional angular galaxy clustering. We test both the standard  $\Lambda$ CDM model and the  $w$ CDM extension. We find that the cosmological and galaxy bias parameters express different degeneracies when inferred from ARF or from angular galaxy clustering. As such, combining both observables breaks these degeneracies and greatly decreases the marginalised uncertainties, by a factor of at least two on most parameters for the  $\Lambda$ CDM and  $w$ CDM model. We find that the ARF combined with angular galaxy clustering are a great probe of dark energy by increasing the figure of merit of the  $w_0$ - $w_a$  parameter set by a factor of more than 10 compared to angular galaxy clustering alone. Finally we compare ARF to the CMB lensing constraints on the galaxy bias parameters. We show that a joint analysis of ARF and angular galaxy clustering improves constraints by  $\sim 40\%$  on galaxy bias compared to a joint analysis of angular galaxy clustering and CMB lensing.

**Key words.** Cosmology: large-scale structure of Universe - observations - cosmological parameters - dark energy

## 1. Introduction

In the coming years, large scale optical and infra-red (IR) surveys will map with unprecedented accuracy our Universe from the present epoch up to when it was roughly one tenth of its current age. A significant part of these surveys will be spectroscopic, e.g., DESI (DESI Collaboration et al. 2016), 4MOST (de Jong & 4MOST Consortium 2015), WEAVE (Bonifacio et al. 2016), or NISP aboard *Euclid* (Laureijs et al. 2011) and will provide us with spectra for large samples of sources. Such spectra will not only enable deep insight on the physics of those objects, but will also yield accurate estimates of their redshift and thus of their distance to the observer. From the cosmological point of view, this will enable a precise (statistical) characterization of the (apparent) spatial distribution of those luminous tracers (via two- or three-point statistics), and this itself should shed precious light on open topics such as, e.g., the nature of dark energy, the possible interplay of dark energy and dark matter, the mass hierarchy of neutrinos, or possible deviations of gravity from General Relativity, to quote a few.

At the same time, a different family of surveys will scan the sky at greater depths with of optical filters, and with exquisite image quality. These photometric experiments build high quality, very large catalogues of sources, with however relatively rough redshift estimations given their moderate number of filters/colours. While mining the faint Universe, this type of surveys will be particularly sensitive, from a cosmological perspective, to the angular clustering of luminous matter, the cosmolog-

ical aspects of gravitational lensing throughout cosmic epochs, the satellite population in halos, and the formation and evolution of the population of galaxy clusters. In this context, the Dark Energy Survey (DES, Abbott et al. 2018) is currently providing state-of-the-art cosmological constraints in the late universe, and these should be further complemented by the Vera Rubin Observatory (LSST, Ivezić et al. 2019), which, at the same time, will also explore the variability of the night sky in a regime of depth and time domain that so far remains practically unexplored.

An intermediate third class of experiments also exists. These are the *spectro-photometric* surveys that conduct standard photometry in a relatively large set (from  $\sim 10$  up to  $\sim 60$ ) of narrow-band optical filters. This strategy combines the indiscriminate character of the photometric surveys with *high* precision redshift estimates ( $\Delta z/(1+z) \sim 10^{-3}$ – $10^{-2}$ ) for a large fraction ( $> 20$ – $30\%$ ) of the detected sources. Given its multi-color character, these surveys are able to provide *pseudo-spectrum / photo-spectrum* in each pixel of the surveyed area. The pioneer example of COMBO-17 has been or is being followed by other efforts such as COSMOS (Scoville et al. 2007), ALHAMBRA (Moles et al. 2008), SHARDS (Pérez-González et al. 2013), PAU (Martí et al. 2014), J-PAS (Benitez et al. 2014), SPHEREx (Doré et al. 2014), and J-PLUS (Cenarro et al. 2019).

In this work, we forecast the power of cosmological analyses in spectroscopic and spectro-photometric surveys. In this type of surveys, it is customary to convert redshift estimates into radial distances under the assumption of a given fiducial cosmological model. Angular and redshift coordinates are thus converted

into the three-dimensional space, where standard 3D clustering analysis techniques are applied.

In our case however, we choose to follow a different strategy. We focus on a new cosmological observable, namely the angular redshift fluctuations (ARF, [Hernández-Monteagudo et al. 2019](#), hereafter HMCMA). Being a 2D observable, the ARF field can easily be cross-correlated with other 2D observables, such as the 2D galaxy density field and the CMB lensing fields. ARF present other interesting features, such as being correlated to the cosmic, radial, peculiar velocity fields (HMCMA, [Chaves-Montero et al. 2019](#)), or being particularly insensitive to additive systematics that remain constant under the redshift shell subject to analysis.

In this paper we apply the Fisher formalism on the angular galaxy clustering, the ARF and the CMB lensing convergence observables, and explore their sensitivity to cosmology in two different observational set ups, mimicking those expected for the DESI and *Euclid* surveys. We consider the CMB lensing convergence field among our observables, since it constitutes an intrinsically different probe whose dependence on the parameters defining the galaxy sample is different from that of angular galaxy clustering and ARF. Our scope is to assess if the ARF field can provide complementary information on the galaxy density field and on the CMB lensing field.

The paper is organized as follows. We introduce the spectroscopic galaxy surveys and CMB experiments that we use in our analysis in Sect. 2. In Sect. 3, we present the angular galaxy clustering, the ARF, and the CMB lensing convergence field. In Sect. 4, we compute the foreseen signal to noise ratios of these combination of probes, while also introducing the covariance among those observables. In Sect. 5, we present the predicted constraints on cosmological parameters in the fiducial  $\Lambda$ CDM scenario. Finally we discuss our findings in Sect. 6 and conclude in Sect. 7.

Throughout this paper, we use the *Planck* 2018 cosmology as our fiducial cosmology. We take the values given in Table 2, in the column 6 (best-fit with BAO), of [Planck Collaboration et al. \(2018a\)](#). We use the following naming conventions: *observable* refers to a spherical 2D field built on measured quantities, such as counts, redshifts, or deflection angles, while *probe* refers to the combination of one or two observables into a summary statistics. In practice, our probes will be the two-point angular power spectra  $C_\ell$ . The redshift due to the Hubble expansion is denoted by  $z$ , while  $z_{\text{obs}}$  is the measured redshift (which includes redshift distortions induced by radial peculiar velocities);  $\Omega_{\text{m},0}$  is the density of matter at  $z = 0$  in units of the critical density and  $H_0$  is the Hubble constant;  $r(z) = \int dz c / H(z)$  is the line of sight comoving distance, and  $dV_\Omega = dV/d\Omega = r^2 dr = r^2(z) c/H(z) dz$  is the comoving volume element per solid angle, with  $d\Omega$  a differential solid angle element. Vectors are in bold font and a hat denotes a unit vector.

## 2. Surveys

Among the wealth of current and upcoming experiments, we choose two representative cases for spectroscopic large scale structures (LSS) surveys, namely the DESI and the *Euclid* experiments. We detail their specifications in Table 1.

On the CMB side, we consider first a *Planck*-like experiment, currently the state-of-the-art data base in terms of multifrequency, full sky CMB data ([Planck Collaboration et al. 2018a](#)). In order to observe the future sensitivity reachable on the smallest angular scales via ground CMB experiments, we also consider the Simons Observatory ([The Simons Observatory Collab-](#)

[oration et al. 2019](#)) and the CMB-Stage 4 ([Abazajian et al. 2019](#)). Both shall cover thousands of square degrees of the southern sky ( $> 40\%$  of the sky) with extremely high sensitivity ( $\leq 2 \mu\text{K arcmin}$ ) and fine angular resolution (at the arcmin level)

### 2.1. The DESI experiment

DESI is a ground-based survey that will cover  $14,000 \text{ deg}^2$  on the sky and will measure the redshift of about 30 million galaxies using optical fibers spectroscopy ([DESI Collaboration et al. 2016](#)). It will target four different classes of galaxies. In this paper, we compute forecasts for the emission line galaxies (ELG) sample which is the largest sample of the survey. It ranges from  $z = 0.6$  up to  $z = 1.6$ . The expected galaxy distribution  $\bar{n}_g(z)$  (see Fig. 1a) and the galaxy bias  $b(z)$  are calibrated based on the DEEP2 survey ([Newman et al. 2013](#)). The (linear) bias of the spatial distribution of this galaxy population with respect to dark matter is a redshift dependent quantity approximated by

$$b_g(z) = 0.84/D(z), \quad (1)$$

with  $D(z)$  denoting the growth factor of linear matter density perturbations.

The DESI experiment successfully achieved the commissioning phase in the first half of 2020 and should start observing the sky in the second half of 2020 for a five year period. The DESI collaboration plans to release annual datasets.

### 2.2. The Euclid spectroscopic survey

The *Euclid* satellite will observe about  $15,000 \text{ deg}^2$  of the extragalactic sky ([Laureijs et al. 2011](#)). The NISP instruments will provide slitless spectroscopy, allowing for precise redshift determinations for about  $1950 \text{ gal deg}^{-2}$ . The spectroscopic survey will target  $H_\alpha$  emission-line galaxies in the redshift range  $0.9 < z < 1.8$ . We assume the model 3 from [Pozzetti et al. \(2016\)](#) for the expected number density of galaxies  $\bar{n}_g(z)$  (see Fig. 1b). For the expected galaxy linear bias, we fit a linearly redshift dependent bias on the values of the Table 3 of [Euclid Collaboration et al. \(2019\)](#), yielding

$$b_g(z) = 0.79 + 0.68z. \quad (2)$$

The *Euclid* satellite will be launched in 2022 and will be operated for six years. The data will be released in three batches: 2, 4 and 7 years after the launch.

### 2.3. Tomography

As already mentioned above, our forecasts are based on a tomographic approach where the entire redshift range covered by a galaxy survey is sliced into different redshift bins. Centred at each of these redshift bins, we consider Gaussian redshift shells of a given width  $\sigma_z$  centered on redshifts  $z_i$ .

$$W_i(z) = \exp\left(-\frac{(z - z_i)^2}{2\sigma_z^2}\right). \quad (3)$$

Provided that a Gaussian shell will be diluting information on radial scales shorter than the Gaussian width, our choice for  $\sigma_z$  is a compromise between maximizing the amount of radial scales under study, and minimizing the impact of non-linear, radial scales in the analysis ([Asorey et al. 2012](#); [Di Dio et al. 2014](#)). In HMCMA, we found that, at  $z \approx 1$ , down to  $\sigma_z = 0.01$  the impact of radial non-linearities was either negligible or easily

Survey	<i>Euclid</i>	DESI
Survey area	15 000 deg <sup>2</sup>	14 000 deg <sup>2</sup>
Redshift estimation	Slitless spectroscopy	Optical fiber spectroscopy
Targets	$H_\alpha$ emission line	[OII] doublet
Redshift range	$0.9 < z < 1.8$	$0.6 < z < 1.6$
Average number of galaxies	1950 gal deg <sup>-2</sup>	1220 gal deg <sup>-2</sup>
Galaxy bias	$b_g(z) = 0.79 + 0.68 z$	$b_g(z) = 0.84/D(z)$
Reference	<a href="#">Euclid Collaboration et al. (2019)</a>	<a href="#">DESI Collaboration et al. (2016)</a>

Table 1: Specifications for the two galaxy surveys under consideration.

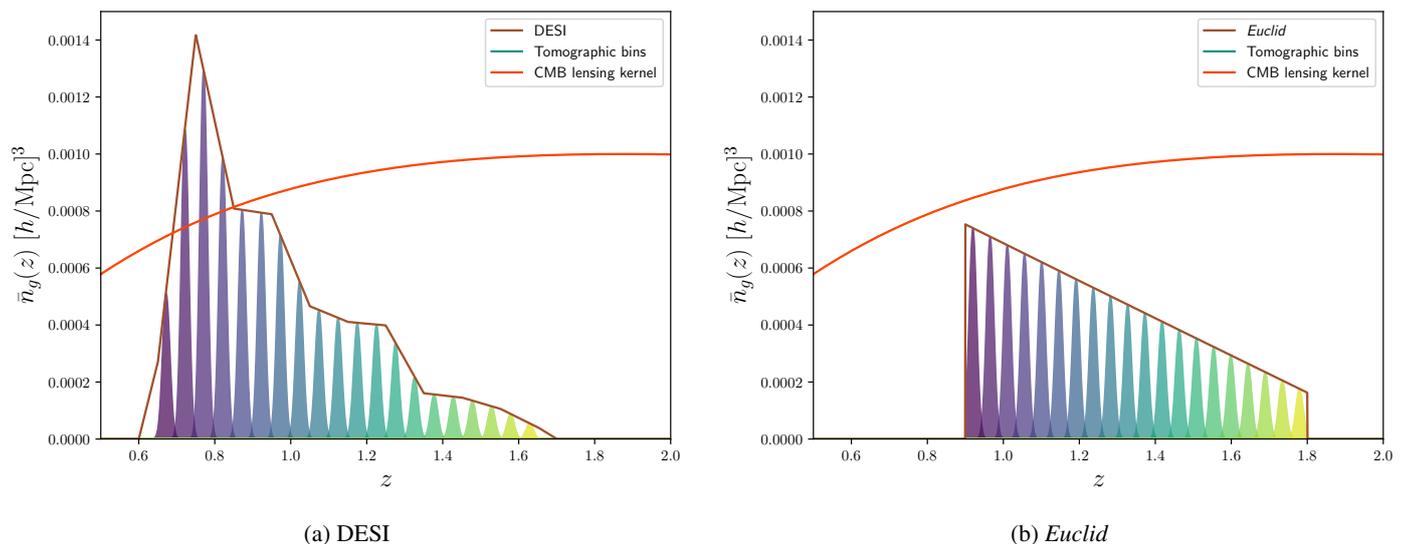


Fig. 1: Galaxy density distribution as a function of redshift for the emission line galaxies (ELG) of DESI (left panel), and the *Euclid* spectroscopic sample (right panel). The filled coloured lines show the Gaussian bins used in our analysis, colour coded as a function of bin index. The orange line shows the CMB lensing efficiency kernel (with arbitrarily normalization).

tractable with a Gaussian kernel describing thermal, stochastic, radial motions. We thus adopt  $\sigma_z = 0.01$  for our forecasts.

As shown in [Asorey et al. \(2012\)](#), the angular galaxy clustering analysis can recover the same amount of information as the 3D analysis when the bin size is comparable to the maximum scale probed by the 3D analysis. This gives  $\sigma_z c/H(z) \simeq 2\pi/k_{\max}$ , so in our case, for  $z = 1$  and  $k_{\max} = 0.2 h \text{ Mpc}^{-1}$  (see Sect. 3), we get  $\sigma_z \simeq 0.01$ , corresponding to our choice of bin size.

For the two galaxy survey under consideration we take the same number of 20 redshift bins, and since the overlap between consecutive bins is not zero, we account for all cross-correlations between shells in the covariance matrix. In this way, redundant information between different shells is fully accounted for. The redshift bins sample the range from  $z = 0.65$  to  $z = 1.65$  for DESI, and from  $z = 0.9$  to  $z = 1.8$  for *Euclid*. These redshift bins are displayed in Figs. 1a and 1b, together with the expected number density of tracers for each survey.

#### 2.4. The Planck experiment

The *Planck* satellite was launched in 2009 and scanned the full sky until 2013 in CMB frequencies. The satellite hosted two instruments, the HFI operating in six frequency bands between 100 GHz and 857 GHz, and the LFI instrument operating in three band between 30 GHz and 77 GHz. The CMB maps were produced by combining these frequencies to remove the contribution from the galaxy and other foregrounds sources. The final

maps have noise of  $27 \mu\text{K arcmin}$ , and an the effective beam with full width at half maximum of 7 arcmin. The final data release of *Planck* was published in [Planck Collaboration et al. \(2018a\)](#).

The CMB lensing field has been estimated with a minimum variance quadratic estimator, combining temperature and polarization data. It is to date the most precise map of the integral of the density of matter on the full extra galactic sky, covering  $\sim 70\%$  of the sky, which allowed to get an estimate of the lensing-potential power spectrum over lensing multipoles  $8 \leq L \leq 400$  ([Planck Collaboration et al. 2018b](#)).

#### 2.5. The Simons Observatory

The Simons Observatory consists of four different telescopes placed in the Atacama Desert in Chile, with the goal of providing an exquisite mapping of the CMB intensity and polarization anisotropies from a few degrees down to arcminutes scales. Three of the telescopes have 0.5 m of aperture, and with an angular resolution close to half a degree, will map 10 % of the sky targeting the moderate to large angular scales. Their primary goal is to measure large-scale polarisation from the background of primordial gravitational waves.

Alongside these small telescopes, one 6 m diameter telescope will observe at 27, 39, 93, 145, 225, and 280 GHz, with an angular resolution close to the arcminute, appropriate to obtain a high resolution map of the lensing potential of the CMB. It is

expected to reach a sensitivity level of  $6 \mu\text{K arcmin}$  on 40 % of the sky.

## 2.6. CMB Stage 4

The CMB Stage 4 (CMB-S4) experiment will be the successor of the Simons Observatory, which will combine resources with the successor of the South Pole telescope and the BICEP/Keck collaborations. Its main scope is to measure the imprint of primordial gravitational waves on the CMB polarization anisotropy, but it will also perform a wide survey with a high resolution which will allow to probe the secondary anisotropies with unprecedented accuracy. Its deep and wide survey will cover  $\sim 60\%$  of the extra-galactic sky and will be conducted over seven years using two 6 m telescopes located in Chile, each equipped with 121,760 detectors distributed over eight frequency bands from 30 GHz to 270 GHz. These observations will provide CMB temperature and polarization maps with a resolution of  $\leq 1.5 \text{ arcmin}$  and with a noise level of  $1 \mu\text{K arcmin}$ . This very-high sensitivity at small scales both in temperature and polarization, on a large fraction of the sky, will allow to get an accurate estimation of the CMB lensing potential.

## 3. Observables

In this paper, we consider three different observables, namely the angular galaxy clustering, the corresponding ARF, and the CMB lensing convergence field. In order to compute the forecasts, we shall restrict to the linear scales, where the cosmological linear theory of perturbations apply. In practice, we ignored all scales above  $k_{\text{max}} = 0.2 h \text{ Mpc}^{-1}$  at all redshifts. This is a conservative approach, as one could consider a scale cutoff which evolves with redshift as in Di Dio et al. (2014). We shall also assume that our observables are Gaussian distributed, and that the information content is completely captured by the two-point momenta, in particular the angular power spectrum, either auto or cross, depending on whether we combine different observables or not.

In what follows, we describe our model of the observables, so that expressions for their angular power spectrum can be derived thereafter.

### 3.1. Galaxy Angular Density Fluctuations

The 3D field of the number density of galaxies is noted as  $n_g(z, \hat{\mathbf{n}})$ , where  $\hat{\mathbf{n}}$  denotes a direction on the sky. The average number density of galaxies at a redshift  $z$  is defined by  $\bar{n}_g(z) = \langle n_g(z, \hat{\mathbf{n}}) \rangle_{\hat{\mathbf{n}}}$ . The 3D field of galaxy density contrast is then given by

$$\delta_g^{\text{3D}}(z, \hat{\mathbf{n}}) = \frac{n_g(z, \hat{\mathbf{n}}) - \bar{n}_g(z)}{\bar{n}_g(z)}. \quad (4)$$

We assume that the galaxy density contrast traces the dark matter density contrast  $\delta_m^{\text{3D}}$  via a scale-independent bias:  $\delta_g^{\text{3D}}(z, \hat{\mathbf{n}}) = b_g(z) \delta_m^{\text{3D}}(z, \hat{\mathbf{n}})$ . This bias depends on the properties of the galaxies used as a tracer for each survey, and are given in Eqs. 1 and 2.

In our analysis, we model the observed redshift of galaxies  $z_{\text{obs}}$  as a 3D field. It is defined as the sum of the redshift induced by the Hubble flow, and the redshift due to the peculiar velocity of galaxies:

$$z_{\text{obs}}(z, \hat{\mathbf{n}}) = z + (1+z) \frac{\mathbf{v}(z, \hat{\mathbf{n}}) \cdot \hat{\mathbf{n}}}{c}, \quad (5)$$

where  $\mathbf{v}$  is the peculiar velocity field of galaxies. We neglect other sources of redshift distortions which are significantly smaller than those considered here (HMCMA).

The angular galaxy clustering field is then modelled by an integral along the line of sight in which, at every redshift  $z$ , only galaxies within the selection function  $W(z_{\text{obs}}; z_i)$  are included:

$$\delta_g^i(\hat{\mathbf{n}}) = \frac{1}{N_g^i} \int_{z=0}^{\infty} dV_{\Omega} \bar{n}_g(z) b_g(z) \delta_m^{\text{3D}}(z, \hat{\mathbf{n}}) W_i[z_{\text{obs}}(z, \hat{\mathbf{n}})], \quad (6)$$

where  $N_g^i = \int_{z=0}^{\infty} dV_{\Omega} \bar{n}_g(z) W_i(z)$  is the average number of galaxies per solid angle, under the  $i$ -th selection function  $W_i$  centred on redshift  $z_i$ , and in practice can be computed from an angular average over the survey's footprint.

We next expand the selection function, retaining only linear terms in density and velocity fluctuations, finding:

$$\delta_g^i(\hat{\mathbf{n}}) \simeq \frac{1}{N_g^i} \int_{z=0}^{\infty} dV_{\Omega} \bar{n}_g(z) W_i(z) \times \left[ b_g(z) \delta_m^{\text{3D}}(z, \hat{\mathbf{n}}) + (1+z) \frac{d \ln W_i}{dz} \frac{\mathbf{v}(z, \hat{\mathbf{n}}) \cdot \hat{\mathbf{n}}}{c} \right], \quad (7)$$

with the derivative  $d \ln W_i / dz = -(z - z_i) / \sigma_z^2$ .

### 3.2. Angular redshift fluctuations

The ARF field represents the spatial variations of the average redshift of galaxies on the sky. The average redshift of galaxies is given by

$$\bar{z} = \frac{1}{N_g^i} \left\langle \int_{z=0}^{\infty} dV_{\Omega} z_{\text{obs}}(z, \hat{\mathbf{n}}) n_g(z, \hat{\mathbf{n}}) W_i[z_{\text{obs}}(z, \hat{\mathbf{n}})] \right\rangle_{\hat{\mathbf{n}}} \\ = \frac{1}{N_g^i} \int_{z=0}^{\infty} dV_{\Omega} z \bar{n}_g(z) W_i(z). \quad (8)$$

We thus define the ARF field as follows,

$$\delta_z^i(\hat{\mathbf{n}}) = \frac{1}{N_g^i} \int_{z=0}^{\infty} dV_{\Omega} (z_{\text{obs}}(z, \hat{\mathbf{n}}) - \bar{z}) \bar{n}_g(z) \times \left[ 1 + b_g(z) \delta_m^{\text{3D}}(z, \hat{\mathbf{n}}) \right] W_i[z_{\text{obs}}(z, \hat{\mathbf{n}})], \quad (9)$$

where we again refer to a redshift bin centred upon  $z_i$ . Expanding the Gaussian selection function at first order and retaining only linear terms in density and velocity, we find:

$$\delta_z^i(\hat{\mathbf{n}}) \simeq \frac{1}{N_g^i} \int_{z=0}^{\infty} dV_{\Omega} \bar{n}_g(z) W_i(z) \left[ (z - \bar{z}) b_g(z) \delta_m^{\text{3D}}(z, \hat{\mathbf{n}}) + (1+z) \frac{\mathbf{v}(z, \hat{\mathbf{n}}) \cdot \hat{\mathbf{n}}}{c} \left( 1 + (z - \bar{z}) \frac{d \ln W_i}{dz} \right) \right]. \quad (10)$$

Note that given the small widths adopted ( $\sigma_z = 0.01$ ), it is safe to assume that the bias  $b(z)$  remains constant within the redshift bin.

### 3.3. CMB lensing

The image of the primary CMB, emitted at the moment of recombination at  $z \simeq 1100$ , is distorted by the gravitational lensing arising as a consequence of the (slightly inhomogeneous) mass distribution between us and the surface of last scattering.

This modify the initial anisotropy pattern and creates statistical anisotropy (see [Lewis & Challinor 2006](#), for a review). Assuming that the primordial CMB is Gaussian and statistically isotropic, we can reconstruct the lensing potential  $\phi$  with the so called quadratic estimator ([Hu & Okamoto 2002](#); [Okamoto & Hu 2003](#)). The lensing potential is linked to the convergence field  $\kappa_{\text{CMB}}$  by

$$\kappa_{\text{CMB}} = -\frac{1}{2}\Delta\phi. \quad (11)$$

This convergence field is directly proportional to the surface mass density along the line of sight. The CMB lensing is as such an unbiased estimation of the distribution of mass. However it is an integrated estimation, whereas galaxy surveys can enable tomographic analyses thanks to redshift measurements.

The CMB lensing has been characterized by the *Planck* CMB survey ([Planck Collaboration et al. 2018b](#)), and by the ACTPol ([Sherwin et al. 2017](#)), SPT-SZ ([Omori et al. 2017](#)) and SPTpol ([Wu et al. 2019](#)) collaborations. Next generation CMB surveys such as Simons Observatory or CMB-S4 will increase the signal to noise ratios at all  $\ell$  by almost one and two orders of magnitude respectively. These new experiments will make the CMB lensing as a sensitive probe of the dark matter distribution, and via cross-correlation studies it will be crucial to constrain the growth rate of structure, the neutrino masses, or the level of primordial non-Gaussianities during the inflationary epoch.

### 3.4. Angular power spectra

Our statistical tool to test cosmological models are the angular two-point power spectra  $C_\ell$  performed over the three fields defined in Sects. 3.1, 3.2 and 3.3. Assuming that the galaxy bias and the growth factors are scale independent, one can show that our (cross and auto) angular power spectra can be expressed as the convolution of two kernels  $\Delta_\ell^A(k)$  and  $\Delta_\ell^B(k)$ , corresponding for the fields A and B (see, e.g. [Huterer et al. 2001](#)):

$$C_\ell^{\text{A,B}} = \frac{2}{\pi} \int dk k^2 P(k) \Delta_\ell^A(k) \Delta_\ell^B(k), \quad (12)$$

where  $P(k)$  is the linear 3D matter power spectrum at  $z = 0$ , function of the wave number  $k$ .

To obtain the theoretical prediction of our angular power spectra, we start from the 2D fields defined in Eqs. 7 and 10. The velocity field is related to the matter density contrast field via the linearized continuity equation  $\partial \delta_m^{\text{3D}} / \partial t + \nabla \mathbf{v} / a = 0$ , with  $a(z)$  the cosmological scale factor,  $a = 1/(1+z)$ . We introduce the linear growth rate

$$f = \frac{d \ln D}{d \ln a} = -(1+z) \frac{1}{D(z)} \frac{dD}{dz}. \quad (13)$$

We assume  $f(z) = \Omega_m(z)^\gamma$ , with  $\gamma = 0.55$  ([Lahav et al. 1991](#); [Linder 2005](#)). The growth factor  $D(z)$  is computed by integrating the growth rate  $f(z)$ .

One can show that the angular galaxy clustering kernel is the sum of two terms, one arising from the density of galaxies and the other from the peculiar line of sight velocities,  $\Delta_\ell^g = \Delta_\ell^g|_\delta + \Delta_\ell^g|_v$  (see e.g. [Padmanabhan et al. 2007](#)):

$$\Delta_\ell^g|_\delta(k) = \frac{1}{N_g^i} \int_{z=0}^{\infty} dV_\Omega \bar{n}_g(z) W_i(z) b_g(z) D(z) j_\ell(kr(z)), \quad (14)$$

$$\Delta_\ell^g|_v(k) = \frac{1}{N_g^i} \int_{z=0}^{\infty} dV_\Omega \bar{n}_g(z) H(z) f(z) D(z) \frac{dW_i}{dz} \frac{j'_\ell(kr(z))}{k}, \quad (15)$$

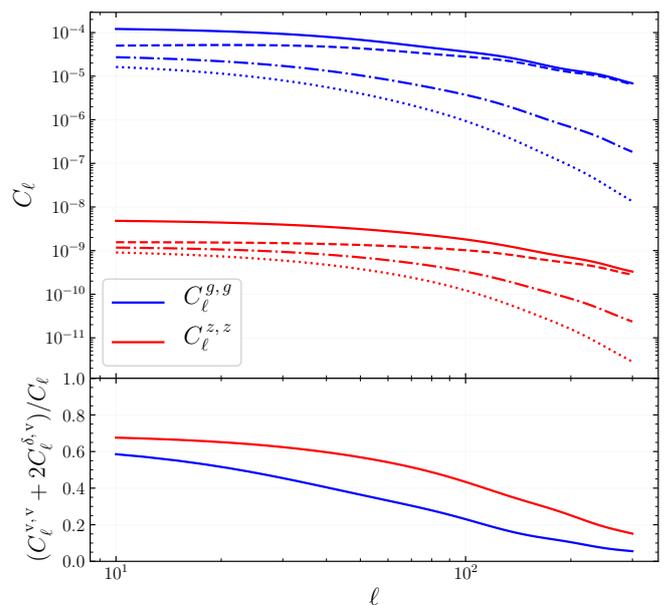


Fig. 2: *Top panel*: Power spectra of for angular galaxy clustering ( $\delta_g$ , in blue) and ARF ( $\delta_z$ , in red), for a Gaussian redshift bin taken in a DESI-like survey. The bin is centered on  $z_i = 0.75$  and has a standard deviation of  $\sigma_z = 0.01$ . The dashed line show the term coming from the density kernel  $C_\ell^{\delta,\delta}$ ; the dotted line show the part coming from the velocity kernel  $C_\ell^{v,v}$  and the dot-dashed line show the cross term  $C_\ell^{\delta,v}$ . The total  $C_\ell$  power spectra (plain lines) correspond to the sum  $C_\ell = C_\ell^{\delta,\delta} + 2C_\ell^{\delta,v} + C_\ell^{v,v}$ . *Bottom panel*: Ratio of the velocity dependence in the power spectrum  $(C_\ell^{v,v} + 2C_\ell^{\delta,v}) / C_\ell$  over the complete power spectrum, for the angular galaxy clustering (blue line) and for the ARF (red line). This figure shows that ARF are more sensitive to the peculiar velocity of galaxies than angular galaxy clustering, for the same redshift shell.

where  $j_\ell(x)$  is the spherical Bessel function of order  $\ell$  and  $j'_\ell(x)$  is its derivative  $j'_\ell(x) \equiv dj_\ell/dx$ .

One can thus write the power spectrum as the sum of the contributions from the density and from the velocity kernels  $C_\ell = C_\ell^{\delta,\delta} + 2C_\ell^{\delta,v} + C_\ell^{v,v}$ .

The ARF kernel can also be separated into two kernels:

$$\Delta_\ell^{z,i}|_\delta(k) = \frac{1}{N_g^i} \int_{z=0}^{\infty} dV_\Omega \bar{n}_g(z) W_i(z) b_g(z) D(z) (z - \bar{z}) j_\ell(kr(z)), \quad (16)$$

$$\Delta_\ell^{z,i}|_v(k) = \frac{1}{N_g^i} \int_{z=0}^{\infty} dV_\Omega \bar{n}_g(z) H(z) f(z) D(z) W_i(z) \times \left[ 1 + (z - \bar{z}) \frac{d \ln W_i}{dz} \right] \frac{j'_\ell(kr(z))}{k}. \quad (17)$$

The kernel function of the CMB lensing convergence field is given by:

$$\Delta_\ell^k(k) = \frac{3\Omega_{m,0}}{2} \left( \frac{H_0}{c} \right)^2 \int_{r=0}^{r_*} dr \frac{r}{a(r)} \frac{r_* - r}{r_*} D(z(r)) j_\ell(kr), \quad (18)$$

where  $r_*$  the comoving distance from the observer to the last scattering surface, and  $a$  is the cosmological scale factor.

The top panel of Fig. 2 shows the angular power spectra of the angular galaxy clustering and ARF, for a Gaussian selection

function of width  $\sigma_z = 0.01$  centered on  $z_i = 0.75$  in a DESI-like survey. We show in the same figure the terms arising from the density fluctuation kernel and the peculiar velocity kernel (c.f. Eqs. 14 to 17). We can see that the peculiar velocity term is relatively more important (compared to the total power spectrum) in the ARF power spectrum than in the angular galaxy clustering power spectrum. To better illustrate this fact, we show in the bottom panel Fig. 2 the ratio of the velocity part of the power spectrum (which is the sum  $C_\ell^{v,v} + 2C_\ell^{\delta,v}$ ) over the total power spectrum for both angular galaxy clustering and ARF. For both fields, the peculiar velocity contribution dominates at low  $\ell$ , while it vanishes to zero for  $\ell > 300$ . At  $\ell = 10$ , the velocity-dependent part in the power spectrum represents around 67% of the total contribution for  $C_\ell^{z,z}$ , while it represents only 58% of  $C_\ell^{g,g}$ . The difference between the two is even more visible at  $\ell = 60$ , where the velocity contribution represents 55% of  $C_\ell^{z,z}$  and only 35% of  $C_\ell^{g,g}$ .

This difference is caused by the intrinsic different nature of the angular galaxy clustering and ARF transfer functions: angular galaxy clustering is sensitive to the average of density and velocity under the Gaussian shell, whereas ARF is sensitive to radial derivatives of those fields. For narrow shells, this makes both fields practically uncorrelated (HMCMA), and given the ratio comparison showed in Fig. 2, one would expect ARF to be more sensitive than angular galaxy clustering to cosmological parameters impacting peculiar velocities.

#### 4. Signal to noise forecasts

We forecast the expected signal to noise ratio (SNR) for different combinations of observables. Our data vector  $\mathbf{D}(\ell)$  contains the auto- and cross-power spectra between the different observables and between the redshift bins. In order to compare several combinations of probes, we define the following data vectors:

$$\begin{aligned} \mathbf{D}_g(\ell) &= (C_l^{g_i, g_j}), \\ \mathbf{D}_z(\ell) &= (C_l^{z_i, z_j}), \\ \mathbf{D}_{g,z}(\ell) &= (C_l^{g_i, g_j}, C_l^{g_i, z_j}, C_l^{z_i, z_j}), \\ \mathbf{D}_{g, \kappa_{\text{CMB}}}(\ell) &= (C_l^{g_i, g_j}, C_l^{g_i, \kappa_{\text{CMB}}}, C_l^{\kappa_{\text{CMB}}, \kappa_{\text{CMB}}}), \\ \mathbf{D}_{g,z, \kappa_{\text{CMB}}}(\ell) &= (C_l^{g_i, g_j}, C_l^{g_i, z_j}, C_l^{g_i, \kappa_{\text{CMB}}}, C_l^{z_i, z_j}, \\ &\quad C_l^{z_i, \kappa_{\text{CMB}}}, C_l^{\kappa_{\text{CMB}}, \kappa_{\text{CMB}}}), \end{aligned} \quad (19)$$

where  $i$  and  $j$  are indexes running over the redshift bins. We perform a tomographic analysis with 20 redshift bins, thus the data vectors containing only the auto-spectra of angular galaxy clustering and ARF ( $\mathbf{D}_g$  and  $\mathbf{D}_z$ ) contain 210  $C_\ell$  each. The data vector containing the cross-correlation  $\mathbf{D}_{g,z}$  has 820  $C_\ell$  and the longest data vector  $\mathbf{D}_{g,z, \kappa_{\text{CMB}}}$  contains 861  $C_\ell$ .

In Fig. 3, we display the correlation matrix for the  $\mathbf{D}_{g,z, \kappa_{\text{CMB}}}(\ell = 10)$  data vector. We clearly see that, in the same redshift bin, angular galaxy clustering and ARF are practically un-correlated (diagonal terms of the top left and lower right blocks close to zero), but that there is some degree of anti-correlation in neighbouring redshift bins. We can also observe that the CMB lensing field is almost uncorrelated with the ARF.

We assume that there is no correlation between different multipoles and that the covariance between the probes is totally captured by a Gaussian covariance. This assumption is exact on large (linear) scales and if the survey covers the full sky. On real

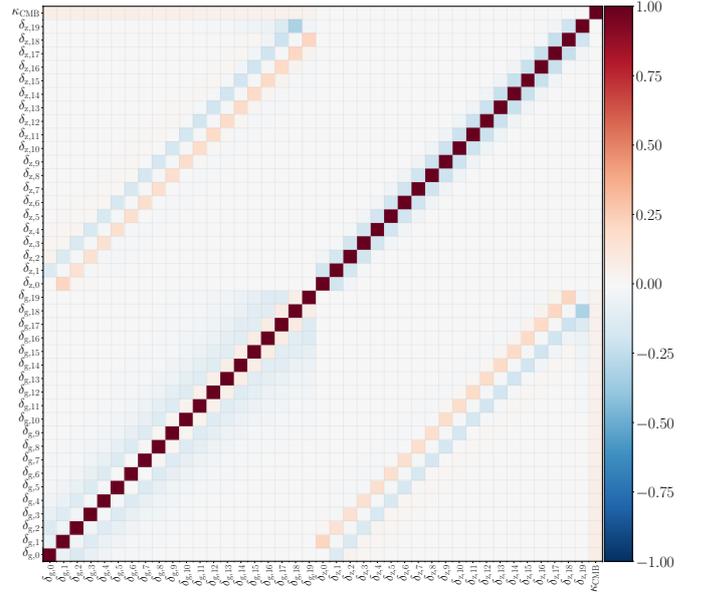


Fig. 3: Correlation matrix between our observables, for the 20 redshift bins in a *Euclid*-like survey, at  $\ell = 10$ . This matrix corresponds to the  $\mathbf{D}_{g,z, \kappa_{\text{CMB}}}(\ell = 10)$  data vector. The value in each pixel corresponds to  $C_\ell^{A,B} / \sqrt{C_\ell^{A,A} C_\ell^{B,B}}$ . All the data vectors considered in Eqs. 19 to 20 are a subset of this matrix. We see that there is no correlation between  $\delta_g$  and  $\delta_z$  inside the same redshift bin (diagonals of the upper left and lower right blocks), and that there are opposite and positive correlation for neighboring bins.

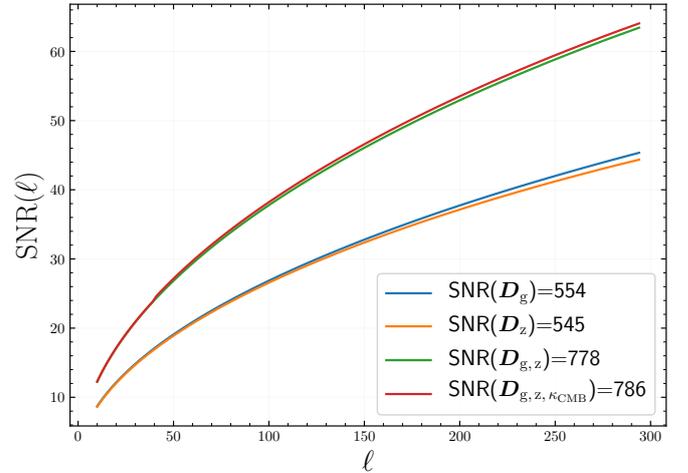


Fig. 4: Signal to noise ratios of angular galaxy clustering ( $\mathbf{D}_g$ ) in blue, ARF ( $\mathbf{D}_z$ ) in orange and the combinations  $\mathbf{D}_{g,z}$  in green and  $\mathbf{D}_{g,z, \kappa_{\text{CMB}}}$  in red. We use 20 tomographic Gaussian bins of width  $\sigma_z = 0.01$  in a *Euclid*-like survey, in combination with a CMB-S4 survey. The total signal to noise ratios for the range of multipoles  $\ell = 10$  to  $\ell = 300$  are shown in the text box on the bottom right.

data, the footprint of the survey and the presence of masked area will create correlations between multipoles. See e.g. Krause & Eifler (2017) or Lacasa (2018) for the inclusion the higher-order (non-Gaussian) terms in the covariance matrix. In this work we neglect these effects.

The SNR of our data vectors as a function of  $\ell$ , taking into account all redshift bins and the correlations between them, are given by

$$\text{SNR}(\mathbf{D}(\ell)) = \sqrt{\mathbf{D}(\ell)^t \text{Cov}_\ell^{-1} \mathbf{D}(\ell)}, \quad (21)$$

and the total SNR are

$$\text{SNR}(\mathbf{D}) = \sqrt{\sum_{\ell=\ell_{\min}}^{\ell_{\max}} [\text{SNR}(\mathbf{D}(\ell))]^2}. \quad (22)$$

Assuming that there is no correlation between different multipoles, we define our Gaussian covariance matrix between our data vectors as in [Hu & Jain \(2004\)](#)

$$\begin{aligned} \text{Cov}_\ell(C_\ell^{\text{A,B}}, C_\ell^{\text{C,D}}) &= \frac{1}{(2\ell + 1) \Delta\ell f_{\text{sky}}} \\ &\times \left[ (C_\ell^{\text{A,C}} + \delta_{\text{A,C}}^{\text{K}} N_\ell^{\text{A}}) (C_\ell^{\text{B,D}} + \delta_{\text{B,D}}^{\text{K}} N_\ell^{\text{B}}) \right. \\ &\left. + (C_\ell^{\text{A,D}} + \delta_{\text{A,D}}^{\text{K}} N_\ell^{\text{A}}) (C_\ell^{\text{B,C}} + \delta_{\text{B,C}}^{\text{K}} N_\ell^{\text{B}}) \right], \end{aligned} \quad (23)$$

with A, B, C, D are the observables  $\{g_i, z_j, \kappa_{\text{CMB}}\}$ ,  $\Delta\ell$  is the width of the multipole bin,  $\delta_{x,y}^{\text{K}}$  is the Kronecker delta,  $N_\ell$  are the probe specific noise power spectra and  $f_{\text{sky}}$  is the sky fraction of the survey considered.

For the sake of simplicity, when combining galaxy surveys with CMB lensing, we always assume a full overlap of the two. As such, the sky fraction  $f_{\text{sky}}$  is always taken to be the one of either DESI or *Euclid*. Even if not accurate, this provides a rough estimate of the available constraining power that the combination of galaxy surveys with CMB lensing will be able to achieve.

We assume that the noise of the angular galaxy clustering and that of the ARF are the shot noise arising from the discrete nature of galaxy surveys. We model it by replacing the power spectrum of dark matter by a Poissonian term,  $P_{\text{shot}}(k, z) = 1/\bar{n}_g(z)$ , into Eq. 12. From this, we can derive the following expressions for the shot noise

$$N_\ell^{g_i, g_j} = \frac{\delta_{i,j}^{\text{K}}}{N_{\text{gal}}^i}, \quad (24)$$

$$N_\ell^{\bar{z}_i, \bar{z}_j} = \frac{\delta_{i,j}^{\text{K}}}{(N_{\text{gal}}^i)^2} \int dV_\Omega \bar{n}_g(z) W(z_i, z) (z - \bar{z}_i)^2, \quad (25)$$

$$N_\ell^{g_i, \bar{z}_j} = \frac{\delta_j^i}{(N_{\text{gal}}^i)^2} \int dV_\Omega \bar{n}_g(z) W(z_i, z) (z - \bar{z}_i) = 0. \quad (26)$$

We can see here that the shot noise cancels out when computing the cross correlation between the angular galaxy density and the ARF fields.

The noise of the CMB lensing field reconstructed from *Planck* is taken from [Planck Collaboration et al. \(2018b\)](#). For the forecasted Simons Observatory CMB lensing noise, we take the publicly available noise curves provided by [The Simons Observatory Collaboration et al. \(2019\)](#)<sup>1</sup>. In practice, we use the noise curves obtained with the internal linear combination (ILC) component separation method, assuming the baseline analysis for a sky fraction of  $f_{\text{sky}} = 0.4$ . For CMB-S4, the lensing noise curve

<sup>1</sup> We use the version 3.1.0 of the noise curves available on [https://github.com/simonsobs/so\\_noise\\_models](https://github.com/simonsobs/so_noise_models)

is taken as the minimum variance N0 bias, which is computed using the code `quicklens`<sup>2</sup>. We assume that CMB-S4 will have a beam size (full width at half maximum) of 1 arcmin, a temperature noise of  $\Delta T = 1 \mu\text{K}$  arcmin and a polarization noise of  $\Delta P = \sqrt{2} \mu\text{K}$  arcmin ([Abazajian et al. 2019](#)).

For both Simons Observatory and CMB-S4,  $\ell = 40$  is the minimum multipole which will be accessible. We assume that these measurements will be combined with the *Planck* lensing signal for lower multipoles. As a result, we use the lensing noise of *Planck* for multipoles below  $\ell = 40$  when forecasting constraints with the Simons Observatory and CMB-S4.

We use the linear matter power spectrum  $P(k)$  computed with the `CLASS` software ([Blas et al. 2011](#)). In order to focus on the linear regime we restrict our analysis to a maximum multipole of  $\ell_{\max} = 300$ . Assuming the Limber approximation  $k = (\ell + 1/2)/\chi(z)$ , this  $\ell_{\max}$  corresponds to  $k = 0.18 h \text{Mpc}^{-1}$  at a redshift of  $z = 0.65$ . Given that we will sample higher redshifts, we will probe larger scales ( $k$  lower than  $0.18 h \text{Mpc}^{-1}$ ). We hence expect little impact from non-linear physics in our observables (these are expected to become relevant on  $k < 0.2 h \text{Mpc}^{-1}$  at  $z = 0$ , and yet shorter at higher redshifts). Again this is a conservative approach as one could consider a multipole cutoff evolving with redshift as in [Di Dio et al. \(2014\)](#). We stress that, in our computations, we do not use the Limber approximation but the full computation of spherical Bessel functions.

Our minimum multipole is chosen to be  $\ell_{\min} = 10$ . To reduce numerical noise and to speed up Fisher matrix computations, we perform a linear binning of the multipoles. In each multipole bin  $[\ell_i, \ell_{i+1}[$ , the binned  $C_\ell$  is the average of the  $C_\ell$ 's that fall in the bin, and the binned multipole is taken as  $\ell = (\ell_i + \ell_{i+1})/2$ . We choose a bin size of  $\Delta\ell = 3$ , which is applied on the full  $\ell$  range. We check that this binning does not impact the constraints from the Fisher matrix by comparing with the case where we do not perform any binning of the multipoles.

We show in Fig. 4 the SNR for a *Euclid*-like survey combined with a CMB-S4 survey, for four combination of probes:  $\mathbf{D}_g$ ,  $\mathbf{D}_z$ ,  $\mathbf{D}_{g,z}$  and  $\mathbf{D}_{g,z,\kappa_{\text{CMB}}}$  following the redshift binning shown in Fig. 1b. The total signal to noise for this four data vectors is respectively 544, 545, 778 and 786. This shows that the tomographic analysis of angular galaxy clustering and ARF have a similarly high SNR. Moreover, the combined analysis  $\mathbf{D}_{g,z}$  brings more information than measuring the angular galaxy clustering alone  $\mathbf{D}_g$ , as the SNR is increased by 40%.

## 5. Fisher forecasts

We use the Fisher formalism to compute *a priori* how well our data vectors defined in Sect. 4 will constrain cosmological parameters in the context of future surveys. As we assumed that there is no correlation between different multipoles, the Fisher matrix can be summed over the multipoles and is given by

$$F_{i,j} = \sum_{\ell_{\min}}^{\ell_{\max}} \frac{\partial \mathbf{D}(\ell)}{\partial \lambda_i} \text{Cov}_\ell^{-1} \frac{\partial \mathbf{D}(\ell)}{\partial \lambda_j}, \quad (27)$$

with  $\mathbf{D}$  one of the data vectors defined in Eqs. 19 to 20,  $\{\lambda_i\}$  is the set of free parameters of our model, and  $\text{Cov}_\ell$  the covariance matrix given in Eq. 23.

The derivatives  $\partial \mathbf{D}(\ell)/\partial \lambda_i$  are computed as the two-point variation with a 1% step around the fiducial value. We have checked that our derivatives are numerically stable when changing the step size.

<sup>2</sup> <https://github.com/dhanson/quicklens>

$\Omega_b$	$\Omega_m$	$n_s$	$h$	$\sigma_8$	$w_0$	$w_a$
0.04897	0.3111	0.9665	0.6766	0.8102	-1	0

Table 2: Fiducial values of the free parameter of our fiducial cosmological model. We first consider only parameters in the standard  $\Lambda$ CDM model, and later we include the  $w_0, w_a$  parameters from the CPL parametrization of Dark Energy.

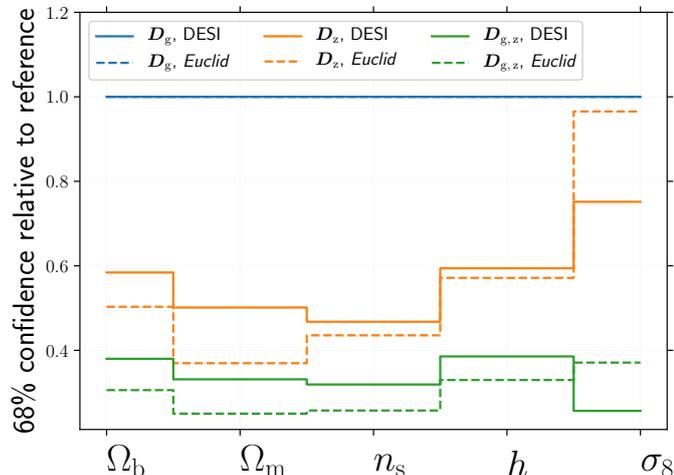


Fig. 5: Ratio of  $1\sigma$  confidence interval relative to the  $1\sigma$  value from angular galaxy clustering ( $D_g$ ) for  $\Lambda$ CDM parameters. Constraints are marginalised over the 20 galaxy bias parameters. Plain lines are for a DESI-like survey while dashed lines are for a *Euclid*-like survey. Blue line shows  $D_g$  (our reference here), orange lines show  $D_z$  and green lines show  $D_{g,z}$ . We see that for most parameters (except  $\sigma_8$ ) confidence intervals shrink by  $\sim 50\%$  when using  $D_z$  instead of  $D_g$ . When using the combination  $D_{g,z}$ ,  $1\sigma$  intervals are shrunk by at least 60% for all parameters.

We compute forecasts for two cosmological models. The first one assumes the standard  $\Lambda$ CDM model, and the parameters we vary are  $\{\Omega_m, \Omega_{\text{baryon}}, \sigma_8, n_s, h\}$ . The fiducial values of these parameters are given by Planck Collaboration et al. (2018a). The second model assumes an evolving dark energy equation of state, with the so-called CPL parametrization (Chevallier & Polarski 2001; Linder 2003):  $w(z) = w_0 + w_a z/(1+z)$ . Our second set of free parameters is then  $\{\Omega_m, \Omega_{\text{baryon}}, \sigma_8, n_s, h, w_0, w_a\}$ . In both cases, we assume a flat universe ( $\Omega_k = 0$ ) with massless neutrinos ( $\sum m_\nu = 0$ ). We show in Tab. 2 the fiducial values of the free parameters.

We also consider a bias parameter assumed constant within each redshift bin, thus adding one free parameter for each redshift shell, over which we marginalise in the Fisher analysis. The fiducial values of the galaxy bias depend on the survey considered and are given in Eqs. 1 and 2. We take the value at  $z_i$ , the center of the Gaussian shell for each bin.

### 5.1. Results for the $\Lambda$ CDM model

The results for the  $\Lambda$ CDM model are summarized in Fig. 5, where we show the ratio of the  $1\sigma$  marginalised uncertainties when including ARF compared to using only angular galaxy clustering, for a DESI-like and a *Euclid*-like surveys. Fig. 6 and

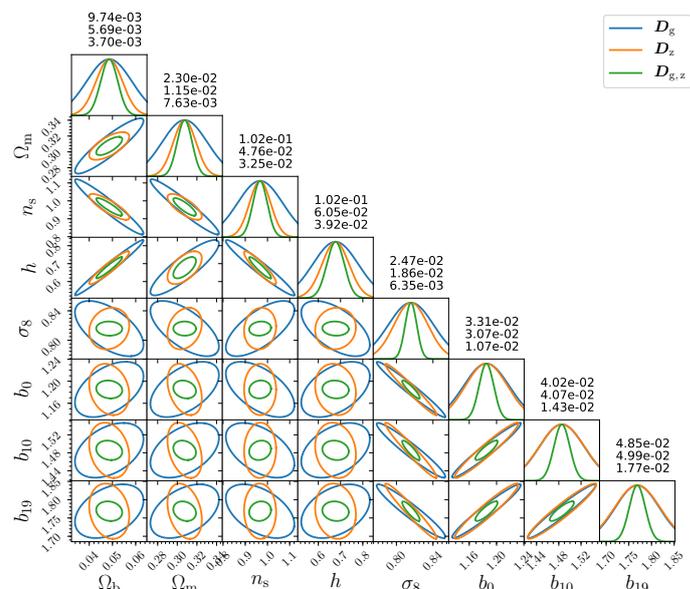


Fig. 6: Foreseen constraints ( $1\sigma$  contours) for a set of five  $\Lambda$ CDM parameters, plus three galaxy bias parameters (out of a total of 20) for a DESI-like survey. We assume 20 tomographic Gaussian bins of size  $\sigma_z = 0.01$ . The blue line are the constraints for angular galaxy clustering alone  $D_g$ , the orange line are for the ARF alone  $D_z$ , and the green line is a joint analysis of both fields  $D_{g,z}$ . The figures above the 1-D PDFs give the marginalised  $1\sigma$  uncertainty of the parameter for each data vector. We show here only 3 galaxy bias parameters even if we marginalised upon the 20 bias parameters.

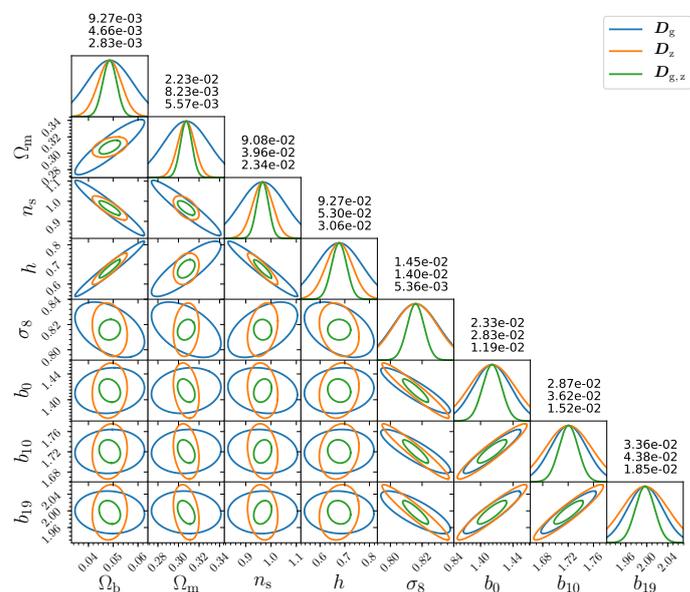


Fig. 7: Same as Fig. 6 for a *Euclid*-like survey.

Fig. 7 show the  $1\sigma$  uncertainty ellipses for the  $\Lambda$ CDM parameters and three out of the 20 galaxy bias parameters for a DESI-like survey and a *Euclid*-like survey respectively. Error ellipses for  $D_g, D_z$ , and  $D_{g,z}$  are given by blue, orange, and green curves, respectively, while marginalised  $1\sigma$  uncertainties for each pa-

parameter are quoted, for these three sets of observables, above the panels containing the one dimensional probability density distributions (PDFs).

For both types of LSS surveys, we can see in Fig. 5 that ARF ( $D_z$ ) are significantly more sensitive than angular galaxy clustering ( $D_g$ ), reducing by a factor two the marginalised uncertainties of all cosmological parameters but  $\sigma_8$ , to which both observables are similarly sensitive. For the combined analysis  $D_{g,z}$ , marginalised uncertainties are reduced by more than 60% for all parameters (including  $\sigma_8$ ), compared to the angular galaxy clustering probe alone  $D_g$ . We find that using ARF in combination with angular galaxy clustering provides almost the same improvement on the constraints on cosmological parameter for both surveys, although the improvement is on average slightly better for our *Euclid*-like survey.

We see in Figs. 6 and 7 that while the degeneracy direction between different cosmological parameter pairs seems very similar for both angular galaxy clustering and ARF, this is again different for  $\sigma_8$ . For  $D_z$  this parameter seems rather independent from other cosmological parameters, while its degeneracy with bias parameters is slightly tilted with respect to that of  $D_g$ . As a consequence the joint  $D_{g,z}$  ellipses show little degeneracy with other parameters, including bias. We also find that the marginalised constraints from both experiments are very close, although the *Euclid*-like experiment provides slightly more sensitive forecasts.

Fig.A.1 in Appendix shows the correlation matrix between our free parameters (including galaxy bias parameters) and illustrates the opposite degeneracies that both  $\sigma_8$  and bias parameters have with the other parameters when comparing ARF and angular galaxy clustering.

Even for those parameters for which both angular galaxy clustering and ARF show a similar direction of degeneracy, the combination of the two observables yields significantly reduced error ellipses. This is mostly due to the lack of correlation between the ARF and angular galaxy clustering for narrow widths used in this work ( $\sigma_z \leq 0.01$ ), as noted in HMCMA and shown here in Fig. 3.

## 5.2. Extension to CPL Dark Energy parametrization

We repeat the analysis detailed above including two new parameters describing the equation of state of dark energy following the CPL parametrization:  $w_0$  and  $w_a$ . We show in Fig. 8 the improvement on the marginalised uncertainties of the ARF with respect to angular galaxy clustering alone. We see that  $D_z$  improves the constraints by 20% to 50% on this set of free  $\Lambda$ CDM parameters, for both surveys. The combined analysis  $D_{g,z}$  reduces the uncertainties by at least 50% and up to 80% for  $\Omega_m$ ,  $\sigma_8$ ,  $w_0$  and  $w_a$ .

The error ellipses are given in Appendix in Fig. B.1 for the DESI-like and *Euclid*-like experiments, displaying a pattern similar to what was found for  $\Lambda$ CDM, together with the correlation matrices (Fig. B.2)

In our idealized case, the combination of ARF with angular galaxy clustering greatly improves the sensitivity of these surveys to dark energy. As shown in Fig. 9, the figure of merit of  $w_0$ - $w_{rma}$  increases by more than a factor of 10 when ARF are combined to angular galaxy clustering. It increases from 17 to 189 for our DESI-like survey and from 19 to 345 for our *Euclid*-like survey.

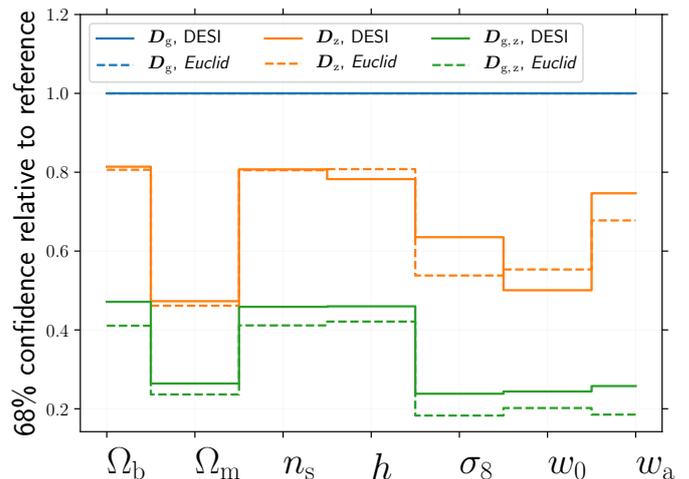


Fig. 8: Ratios of the  $1\sigma$  marginalised uncertainties relative to the  $1\sigma$  marginalised uncertainty for  $D_g$ . We assume  $\Lambda$ CDM model and marginalise on 20 galaxy bias parameters (one for each redshift bin). Orange lines show the ratio for  $D_z$  and green lines show the ratio for  $D_{g,z}$ . Solid lines are for a DESI-like survey while dashed lines are for a *Euclid*-like survey. We see that  $D_z$  improves constraints by up to 50% compared to  $D_g$  and the combined analysis  $D_{g,z}$  improves constraints by up to 80%.

## 5.3. Combining with CMB lensing

We show in Fig. 10 and Fig. 11 the improvements on the constraints of the  $\Lambda$ CDM and  $w$ CDM parameters for a *Euclid*-like survey, when combined with CMB lensing from *Planck*, Simons Observatory and CMB-S4, marginalised over the galaxy bias parameters. We see that including CMB lensing from *Planck* improves the constraints by maximum of 10% in both cosmologies. The improvement is more significant when combining with Simons Observatory or CMB-S4. For the Simons Observatory and CMB-S4, in the  $\Lambda$ CDM model, marginalised uncertainties on  $\Omega_m$  and  $\sigma_8$  are decreased by up to 30%. Other parameters are improved by 5% to 10%. For the  $w$ CDM model, the improvement is of  $\sim 15\%$  for most parameters, with the most significant for  $\Omega_m$  and  $w_a$ , with uncertainties decreased by up to 30%. We see that the combination with CMB lensing help decreasing uncertainties on the  $w$ CDM cosmology.

Since the CMB lensing is an unbiased probe of the distribution of matter, one of the main interests of combining it with galaxy surveys is to produce tight constraints on the galaxy bias parameter. We show in Fig. 12 the  $1\sigma$  marginalised uncertainties on the galaxy bias parameters for each of the 20 redshift bins in a *Euclid*-like survey combined with CMB-S4 lensing, for the  $\Lambda$ CDM model. We compare the constraints obtained for angular galaxy clustering alone ( $D_g$ ), with the ones obtained when combined with CMB lensing ( $D_{g,\kappa_{\text{CMB}}}$ ), with ARF ( $D_{g,z}$ ), and the full combination ( $D_{g,z,\kappa_{\text{CMB}}}$ ).

We see that the combination of angular galaxy clustering with ARF provides better constraints on the galaxy bias than the combination with CMB lensing. For instance, at a redshift of 1.06, the marginalised uncertainties for the galaxy bias parameter  $b_3$  is of 0.025 for the angular galaxy clustering, it decreases to 0.020 when combined with CMB lensing, and down to 0.013 when combined with ARF. The combination of the three results in marginalised uncertainties of 0.08. We can see that the CMB lensing improves constraints by  $\sim 20\%$  only, while ARF im-

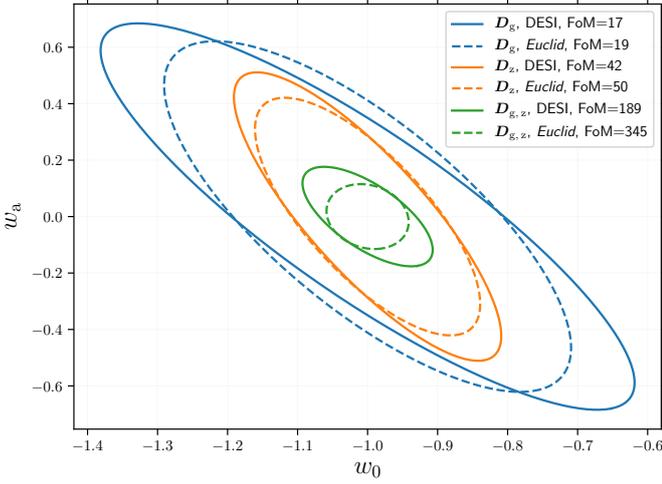


Fig. 9: Marginalised constraints ( $1\sigma$  contours) on the dark energy equation of state parameters for the DESI-like (solid lines) and for the *Euclid*-like (dashed lines) surveys, assuming 20 tomographic Gaussian bins of size  $\sigma_z = 0.01$ . The blue lines are the constraints for angular galaxy clustering alone, the orange lines are for ARF alone, and the green lines are a joint analysis of both fields,  $D_{g,z}$ . These contours are marginalised over the set of cosmological parameters as before, and over the galaxy bias in the 20 redshift bins. We display the figure of merit (FoM) of this pair of parameters in the upper right box for each combination of observables and for each survey.

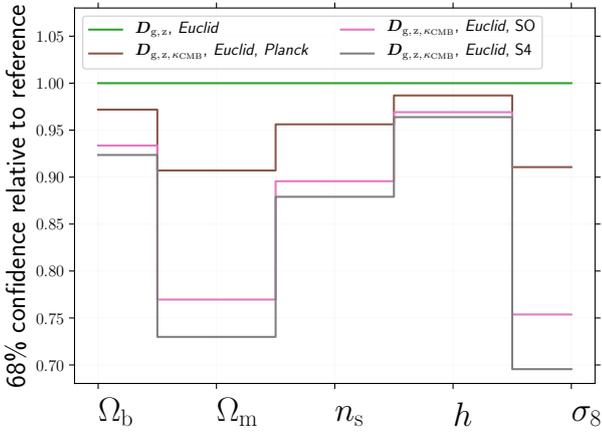


Fig. 10: Ratio of the  $1\sigma$  constraints for  $\Lambda$ CDM parameters from  $D_{g,z,\kappa\text{CMB}}$  over the  $1\sigma$  constraints from  $D_{g,z}$  for the *Euclid*-like spectroscopic survey. We show combinations with CMB lensing from *Planck* (brown), Simons Observatory (pink) and CMB Stage-4 (grey). Constraints are marginalised over the 20 galaxy bias parameters.

proves constraints by  $\sim 50\%$  (a factor 2 improvement). We argue that this is due to the importance of the velocity term in the ARF kernel (see Fig 2), which does not depend on galaxy bias as it is sensitive to the full matter distribution.

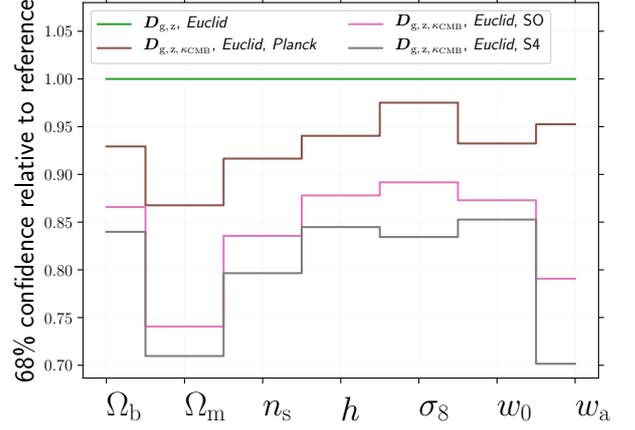


Fig. 11: Same as Fig. 10 for  $w$ CDM parameters.

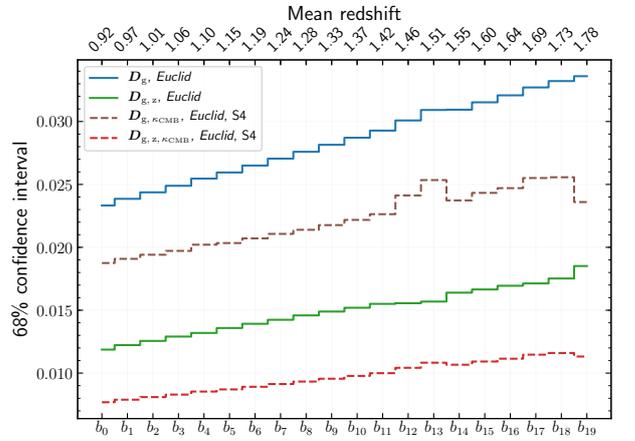


Fig. 12: Marginalised  $1\sigma$  confidence values for the galaxy bias parameters, with a *Euclid*-like survey alone (plain lines) or in combination with CMB-S4 lensing survey (dashed lines). We marginalised on the 5 free parameters of the  $\Lambda$ CDM model. We show constraints with angular galaxy clustering (blue and brown lines) and in combination with ARF (green and red lines). The mean redshift of each shell is shown at the top. We see that the ARF combined with angular galaxy clustering  $D_{g,z}$  provides better constraints on galaxy bias than the combination of angular galaxy clustering with CMB lensing  $D_{g,\kappa\text{CMB}}$ .

## 6. Discussion

One could argue several reasons why angular observables might be preferred over standard 3D ones. Probably the main one is the lack of assumption of any fiducial cosmological model to analyse the data. This means that angular observables may be *directly* compared with theoretical predictions without any intermediate data manipulations that hinge on an assumption whose implications in the analysis may not always be clear. Moreover, this type of angular analysis are conducted tomographically in moderately narrow redshift shells, thus avoiding the assumption that the universe remains effectively frozen in relatively long time spans, as it may occur in 3D clustering analysis where an *effective* redshift must be defined for the entire volume under analysis (see, e.g., Cuesta et al. 2016). Asorey et al. (2012) and Di Dio et al. (2014) have shown that when using a large number of narrow redshift

slices, a 2D clustering analysis can produce the same constraints on cosmological parameters than a 3D clustering analysis, provided that the width of the redshift slices is comparable to the minimum scale probed in the 3D analysis. By including the redshift information in a 2D field, the ARF observable keeps some information about the distribution of galaxies along the line of sight, which normally disappear when projecting the 3D galaxy density field on a 2D observable. As we have shown, ARF are improving the usual 2D galaxy clustering analysis.

Another major interest of using angular observables is that they can easily be cross-correlated with other 2D observables. Indeed the combination of 3D probes with 2D probes is not straightforward, especially when one has to properly take into account the covariances between them (see e.g. Passaglia et al. 2017; Camera et al. 2018). In this work we used the CMB lensing field and its cross-correlation with our tomographic analysis of angular galaxy clustering and ARF. We have shown that this cross-correlations improve the constraints, especially on the galaxy bias. Chaves-Montero et al. (2019) have shown that the cross-correlation of the ARF field with the CMB temperature field can detect the kinematic Sunyaev-Zel'dovich (kSZ) effect at the  $10\sigma$  level.

The point of this paper is not a detailed comparison between 2D and 3D clustering analysis, but rather an exploration of the added value of including ARF in cosmological studies of the large scale structures, on top of the traditional angular galaxy clustering. By its intrinsic different sensitivity to the cosmic density and velocity fields under the redshift shells, the ARF change the degeneracies between cosmological parameters, especially with respect to  $\sigma_8$  and the galaxy bias, compared to the angular galaxy clustering. This is due, as claimed in HMCMA, to the fact that angular galaxy clustering is sensitive to the first moment (the average) of matter density and velocity under the redshift shells, whereas ARF are sensitive to the second moment (the variance) of matter density and velocity along the line of sight, inside these redshift shells. Moreover, we have shown that the ARF and the angular galaxy clustering inside the same tomographic redshift bin are almost uncorrelated. Due to this absence of correlation, by combining both we are able to break degeneracies and give tighter constraints on all the cosmological parameters we have considered.

The results we obtained in our work can be considered as an optimistic setting for both galaxy and CMB surveys. We restricted our analysis to the linear regime and we did not include any systematic effects which could impact our results and worsen the constraints. It was found in Hernández-Monteagudo et al. (2020) that the impact of non-linear physics is more severe in angular galaxy clustering than in ARF. They found that a linear bias was sufficient to describe the ARF on scales larger than  $60 h^{-1}$  Mpc, while it was not the case for angular galaxy clustering. Indeed, ARF are built upon the average observed redshift along the line of sight in a redshift selection function. This is intrinsically different to counting the number of galaxies in a given region in the universe, and consequently systematics and non-linearities are affecting differently each observable. In future works we plan to address systematics and non-linearities, aiming to model more realistic settings. We expect that the impact of both systematics and non-linearities will depend on the survey and on the targeted galaxy sample, as ongoing work on existing galaxy surveys is indicating.

We do not provide a detailed comparison with the forecasted constraints of the *Euclid* survey published in Euclid Collaboration et al. (2019). Indeed our analysis consider a simplistic, linear model of the galaxy clustering. In this context, our findings

indicate that ARF brings significant cosmological information on top of the traditional angular galaxy clustering. At best, our results with the angular galaxy clustering probe ( $\mathbf{D}_g$ ) could be compared with the linear setting shown in the Table 9 of Euclid Collaboration et al. (2019) (first line). In that case, their probe is the 3D linear galaxy power spectrum, with a cutoff value at  $k_{\max} = 0.25 h \text{ Mpc}^{-1}$ , in four different redshift bins. Their Fisher analysis account for more parameters describing the anisotropies in the power spectrum and the shot noise residuals. This 3D probe is intrinsically different to the (2D) angular power spectrum tomography used in our work, in 20 Gaussian bins, for which we limit to  $k_{\max} = 0.20 h \text{ Mpc}^{-1}$ . Our forecasts with  $\mathbf{D}_g$  for the errors on some parameters are tighter than theirs (by a factor of  $\sim 2$  for  $\sigma_8$ ), while for others we find the opposite situation (e.g., the reduced Hubble parameter  $h$ , whose uncertainty in Euclid Collaboration et al. 2019 is roughly one third of ours).

## 7. Conclusion

We showed that the ARF are a promising cosmological observable for next generation spectroscopic surveys. We found that for our choice of binning, the tomographic analysis of ARF retrieves more information than the tomographic analysis of the angular galaxy clustering. We showed that the joint analysis of both fields helps in breaking degeneracies between cosmological parameters, due to their lack of correlation and their different sensitivities to cosmology. The improvement appears to be particularly significant for the  $w_0$ - $w_a$  parameters was increased by a factor of more than ten when combining angular galaxy clustering with ARF.

Finally, we have seen that combining angular galaxy clustering with ARF provides tighter constraints on the galaxy bias parameters compared to the combination of angular galaxy clustering with CMB lensing. This shows that ARF are a very powerful probe of the distribution of matter, as it allows to break the degeneracy between  $\sigma_8$  and the galaxy bias. For future galaxy surveys, errors on the cosmological figure of merit will be dominated by systematic uncertainties and non-linearities, and ARF might provide a novel, complementary view on those issues.

In our analysis we did not consider massive neutrinos. As the growth rate is particularly sensitive to them, we expect ARF to be a powerful tool to constrain the mass of neutrinos. We defer this detailed analysis to an upcoming work.

Simultaneously from the LSS and CMB fronts, the coincidence in the acquisition of data sets of tremendous quality and huge size should enable the combination of standard analyses with new, alternative ones like the one introduced in this paper. The combination of techniques and observables should work jointly in the efforts of identifying and mitigating systematics, and pushing our knowledge of cosmological physics to its limits.

*Acknowledgements.* The authors acknowledge useful discussions with G. Aricò, G. Hurier, and J. Kuruville. LL acknowledges financial support from CNES's funding of the Euclid project. C.H.-M. acknowledges the support of the Spanish Ministry of Science and Innovation through project PGC2018-097585-B-C21. NA acknowledges support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement ERC-2015-AdG 69556.

## References

Abazajian, K., Addison, G., Adshead, P., et al. 2019, arXiv e-prints, arXiv:1907.04473

- Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018, *Phys. Rev. D*, 98, 043526
- Asorey, J., Crocce, M., Gaztañaga, E., & Lewis, A. 2012, *MNRAS*, 427, 1891
- Benitez, N., Dupke, R., Moles, M., et al. 2014, arXiv e-prints, arXiv:1403.5237
- Blas, D., Lesgourgues, J., & Tram, T. 2011, *J. Cosmology Astropart. Phys.*, 2011, 034
- Bonifacio, P., Dalton, G., Trager, S., et al. 2016, in *SF2A-2016: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, ed. C. Rey, J. Richard, L. Cambrésy, M. Deleuil, E. Pécontal, L. Tresse, & I. Vauglin, 267–270
- Camera, S., Fonseca, J., Maartens, R., & Santos, M. G. 2018, *MNRAS*, 481, 1251
- Cenarro, A. J., Moles, M., Cristóbal-Hornillos, D., et al. 2019, *A&A*, 622, A176
- Chaves-Montero, J., Hernández-Monteagudo, C., Angulo, R. E., & Emberson, J. D. 2019, arXiv e-prints, arXiv:1911.10690
- Chevallier, M. & Polarski, D. 2001, *International Journal of Modern Physics D*, 10, 213
- Cuesta, A. J., Vargas-Magaña, M., Beutler, F., et al. 2016, *MNRAS*, 457, 1770
- de Jong, R. S. & 4MOST Consortium. 2015, in *IAU General Assembly*, Vol. 29, 2255843
- DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv e-prints, arXiv:1611.00036
- Di Dio, E., Montanari, F., Durrer, R., & Lesgourgues, J. 2014, *J. Cosmology Astropart. Phys.*, 2014, 042
- Doré, O., Bock, J., Ashby, M., et al. 2014, arXiv e-prints, arXiv:1412.4872
- Euclid Collaboration, Blanchard, A., Camera, S., et al. 2019, arXiv e-prints, arXiv:1910.09273
- Hernández-Monteagudo, C., Chaves-Montero, J., & Angulo, R. E. 2019, arXiv e-prints, arXiv:1911.12056
- Hernández-Monteagudo, C., Chaves-Montero, J., Angulo, R. E., & Ariccò, G. 2020, arXiv e-prints, arXiv:2005.06568
- Hu, W. & Jain, B. 2004, *Phys. Rev. D*, 70, 043009
- Hu, W. & Okamoto, T. 2002, *ApJ*, 574, 566
- Huterer, D., Knox, L., & Nichol, R. C. 2001, *ApJ*, 555, 547
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, 873, 111
- Krause, E. & Eifler, T. 2017, *MNRAS*, 470, 2100
- Lacasa, F. 2018, *A&A*, 615, A1
- Lahav, O., Lilje, P. B., Primack, J. R., & Rees, M. J. 1991, *MNRAS*, 251, 128
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193
- Lewis, A. & Challinor, A. 2006, *Phys. Rep.*, 429, 1
- Linder, E. V. 2003, *Phys. Rev. Lett.*, 90, 091301
- Linder, E. V. 2005, *Phys. Rev. D*, 72, 043529
- Martí, P., Miquel, R., Castander, F. J., et al. 2014, *MNRAS*, 442, 92
- Moles, M., Benítez, N., Aguerra, J. A. L., et al. 2008, *AJ*, 136, 1325
- Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, *ApJS*, 208, 5
- Okamoto, T. & Hu, W. 2003, *Phys. Rev. D*, 67, 083002
- Omori, Y., Chown, R., Simard, G., et al. 2017, *ApJ*, 849, 124
- Padmanabhan, N., Schlegel, D. J., Seljak, U., et al. 2007, *MNRAS*, 378, 852
- Passaglia, S., Manzotti, A., & Dodelson, S. 2017, *Phys. Rev. D*, 95, 123508
- Pérez-González, P. G., Cava, A., Barro, G., et al. 2013, *ApJ*, 762, 46
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018a, arXiv e-prints, arXiv:1807.06209
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018b, arXiv e-prints, arXiv:1807.06210
- Pozzetti, L., Hirata, C. M., Geach, J. E., et al. 2016, *A&A*, 590, A3
- Scoville, N., Aussel, H., Benson, A., et al. 2007, *ApJS*, 172, 150
- Sherwin, B. D., van Engelen, A., Sehgal, N., et al. 2017, *Phys. Rev. D*, 95, 123529
- The Simons Observatory Collaboration, Ade, P., Aguirre, J., et al. 2019, *J. Cosmology Astropart. Phys.*, 2, 056
- Wu, W. L. K., Mocalu, L. M., Ade, P. A. R., et al. 2019, *ApJ*, 884, 70

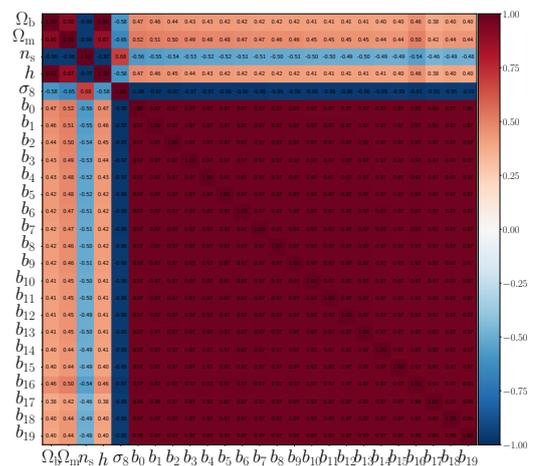
## Appendix A: Correlation matrices

The correlation matrices in Fig. A.1 provide an alternative view of our results. It shows the correlation matrices for the 5  $\Lambda$ CDM parameters and the 20 galaxy bias parameters for a DESI-like survey. We see the opposite correlation of the cosmological parameters  $\Omega_b$ ,  $\Omega_m$ ,  $n_s$  and  $h$  with  $\sigma_8$  for angular galaxy clustering and ARF. This opposite correlation is mirrored in the correlations of those three cosmological parameters with galaxy bias parameters. This is expected as  $\sigma_8$  and bias are tightly correlated. The different nature of the correlation of  $\sigma_8$  and bias with the other cosmological parameters for angular galaxy clustering and ARF is critical for (partially) breaking degeneracies when combining angular galaxy clustering with ARF.

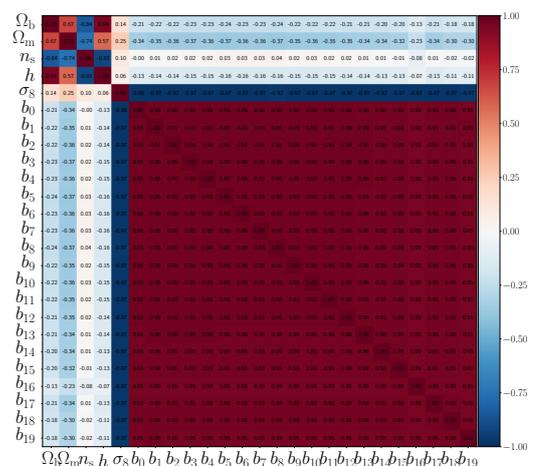
## Appendix B: Results for $w$ CDM model

We show the ellipses obtained with our Fisher analysis for the  $w$ CDM model in Fig. B.1, for a DESI-like and a *Euclid*-like survey. For many parameter pairs, the degeneracy direction (or ellipse orientation) for angular galaxy clustering and ARF are similar, although the resulting error ellipse in the joint  $D_{g,z}$  probe shrink very significantly in all cases. As a result, foreseen uncertainties in the parameters are divided by a factor of at least two for all parameters.

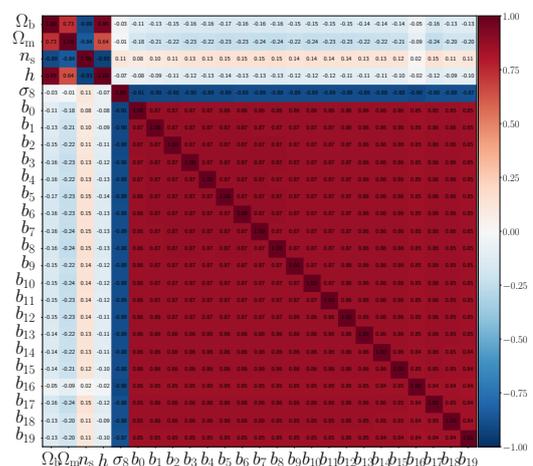
We show in Fig. B.2 the correlation matrices for the  $w$ CDM model. It turns out that for  $D_z$  the new parameters  $w_0$ ,  $w_a$ , together with  $\sigma_8$ , constitute an almost separate (or largely un-correlated) box with respect to all other parameters (see Fig. B.2b). This does not seem to be the situation for angular galaxy clustering  $D_g$  (Fig. B.2a), although this character remains (to great extent) for the joint observable set ( $D_{g,z}$ , Fig. B.2c).



(a)  $D_g$

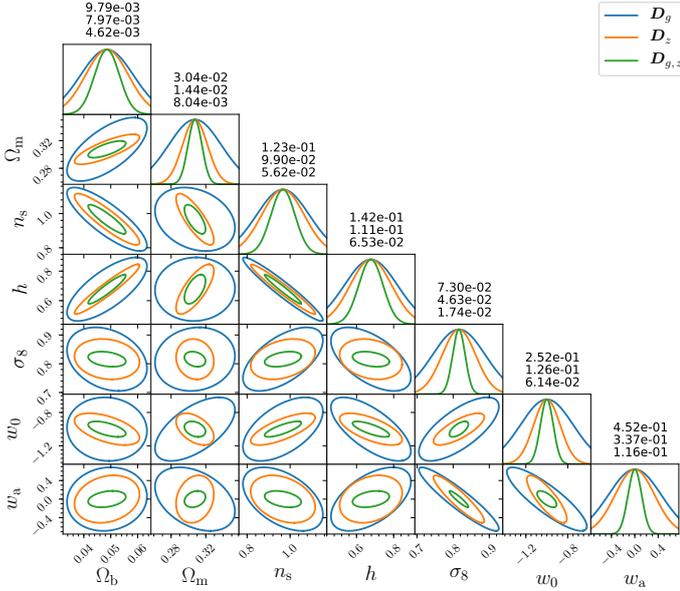


(b)  $D_z$

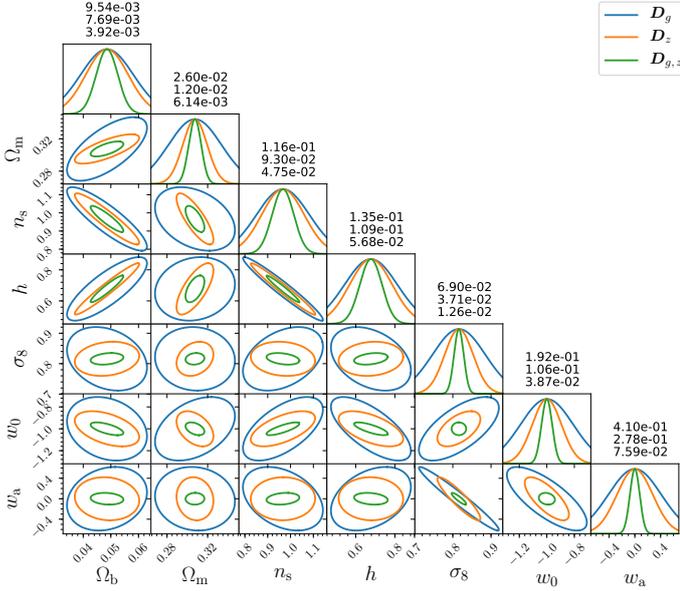


(c)  $D_{g,z}$

Fig. A.1: Correlation between parameters of the  $\Lambda$ CDM model, for a DESI like survey. The top panel is for the angular galaxy clustering alone, the central panel is for ARF alone, and the bottom panel is when combining both observables. We see that the angular galaxy clustering and ARF have opposite correlation coefficients between cosmological parameters and the galaxy bias. The combination of both helps greatly in breaking degeneracies with the galaxy bias.



(a) DESI



(b) Euclid

Fig. B.1: Foreseen constraints ( $1\sigma$  contours) in the CPL cosmological extension  $w$ CDM for a DESI-like survey (top) and a *Euclid*-like survey (bottom), assuming 20 tomographic Gaussian bins of size  $\sigma_z = 0.01$ . The blue line are the constraints for angular galaxy clustering alone ( $D_g$ ), the orange line are for the ARF alone ( $D_z$ ), and the green line is a joint analysis of both fields ( $D_{g,z}$ ). These contours are marginalised over the galaxy bias in the 20 redshift bins. The numbers above the parameter PDFs give the marginalised  $1\sigma$  uncertainty of each parameter for each data-vector.

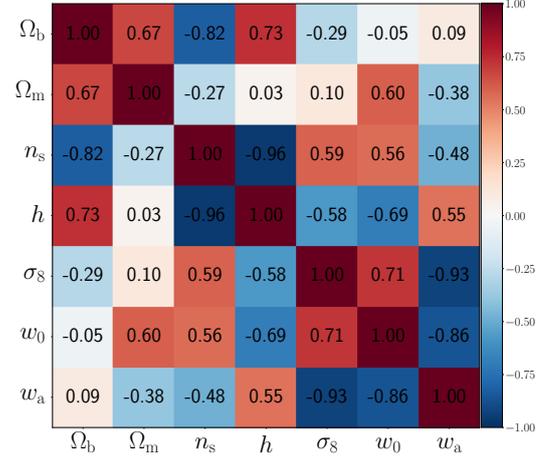
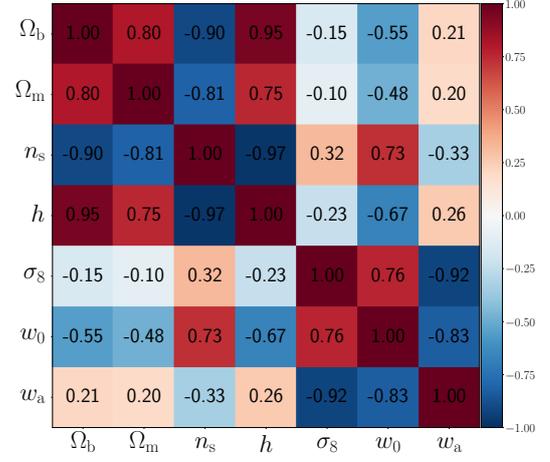
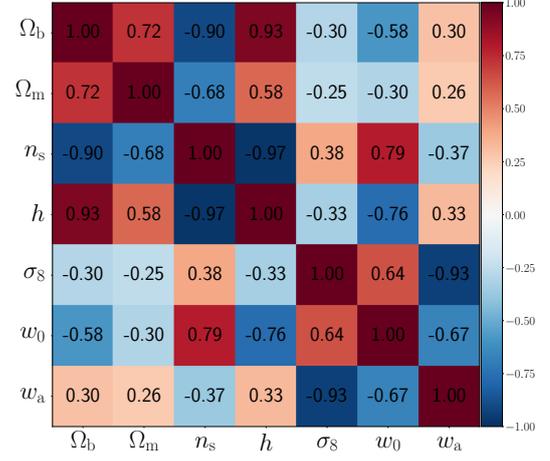
(a)  $D_g$ (b)  $D_z$ (c)  $D_{g,z}$ 

Fig. B.2: Correlation between parameters of the  $w$ CDM model, marginalised over the galaxy bias for a DESI-like survey. The top panel is for the angular galaxy clustering ( $D_g$ ), the central panel refers to ARF alone ( $D_z$ ), and the bottom panel is when combining both observables ( $D_{g,z}$ ). We find that  $D_z$  show different correlations compared to  $D_g$ . The combination of both greatly helps in breaking degeneracies as we can see in panel (c).



# Bibliography

- K. Abazajian, G. Addison, P. Adshead, Z. Ahmed, S. W. Allen, D. Alonso, M. Alvarez, A. Anderson, K. S. Arnold, C. Baccigalupi, et al. CMB-S4 Science Case, Reference Design, and Project Plan. *arXiv e-prints*, art. arXiv:1907.04473, July 2019.
- T. M. C. Abbott, F. B. Abdalla, A. Alarcon, J. Aleksić, S. Allam, S. Allen, A. Amara, J. Annis, J. Asorey, S. Avila, et al. Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing. *Phys. Rev. D*, 98(4):043526, Aug. 2018. doi: 10.1103/PhysRevD.98.043526.
- T. M. C. Abbott, F. B. Abdalla, A. Alarcon, S. Allam, J. Annis, S. Avila, K. Aylor, M. Banerji, N. Banik, E. J. Baxter, et al. Dark Energy Survey year 1 results: Joint analysis of galaxy clustering, galaxy lensing, and CMB lensing two-point functions. *Phys. Rev. D*, 100(2):023541, July 2019. doi: 10.1103/PhysRevD.100.023541.
- N. Aghanim, F. X. Desert, J. L. Puget, and R. Gispert. Ionization by early quasars and cosmic microwave background anisotropies. *A&A*, 311:1–11, July 1996.
- N. Aghanim, S. Majumdar, and J. Silk. Secondary anisotropies of the CMB. *Reports on Progress in Physics*, 71(6):066902, June 2008. doi: 10.1088/0034-4885/71/6/066902.
- Z. Ahmed, M. Amiri, S. J. Benton, J. J. Bock, R. Bowens-Rubin, I. Buder, E. Bullock, J. Connors, J. P. Filippini, J. A. Grayson, et al. BICEP3: a 95GHz refracting telescope for degree-scale CMB polarization. In *Proc. SPIE*, volume 9153 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 91531N, Aug. 2014. doi: 10.1117/12.2057224.
- S. Aiola, E. Calabrese, L. Maurin, S. Naess, B. L. Schmitt, M. H. Abitbol, G. E. Addison, P. A. R. Ade, D. Alonso, M. Amiri, et al. The Atacama Cosmology Telescope: DR4 Maps and Cosmological Parameters. *arXiv e-prints*, art. arXiv:2007.07288, July 2020.
- S. W. Allen, A. E. Evrard, and A. B. Mantz. Cosmological Parameters from Observations of Galaxy Clusters. *ARA&A*, 49(1):409–470, Sept. 2011. doi: 10.1146/annurev-astro-081710-102514.
- L. Amendola, S. Appleby, A. Avgoustidis, D. Bacon, T. Baker, M. Baldi, N. Bartolo, A. Blanchard, C. Bonvin, S. Borgani, et al. Cosmology and fundamental physics with the Euclid satellite. *Living Reviews in Relativity*, 21(1):2, Apr. 2018. doi: 10.1007/s41114-017-0010-3.
- R. E. Angulo and S. D. M. White. One simulation to fit them all - changing the background parameters of a cosmological N-body simulation. *MNRAS*, 405:143–154, June 2010. doi: 10.1111/j.1365-2966.2010.16459.x.

## BIBLIOGRAPHY

---

- M. A. Aragon-Calvo, M. C. Neyrinck, and J. Silk. How Cosmic Web Detachment Drives Galaxy Quenching. *ArXiv e-prints*, July 2016.
- J. Asorey, M. Crocce, E. Gaztañaga, and A. Lewis. Recovering 3D clustering information with angular correlations. *MNRAS*, 427(3):1891–1902, Dec. 2012. doi: 10.1111/j.1365-2966.2012.21972.x.
- É. Aubourg, S. Bailey, J. E. Bautista, F. Beutler, V. Bhardwaj, D. Bizyaev, M. Blanton, M. Blomqvist, A. S. Bolton, J. Bovy, et al. Cosmological implications of baryon acoustic oscillation measurements. *Phys. Rev. D*, 92(12):123516, Dec. 2015. doi: 10.1103/PhysRevD.92.123516.
- J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay. The Statistics of Peaks of Gaussian Random Fields. *ApJ*, 304:15, May 1986. doi: 10.1086/164143.
- J. G. Bartlett and P. Amram. Galactic Free-free Emission and  $H\alpha$ . *arXiv e-prints*, art. astro-ph/9804330, Apr. 1998.
- P. Behroozi, R. Wechsler, A. Hearin, and C. Conroy. UniverseMachine: The Correlation between Galaxy Growth and Dark Matter Halo Assembly from  $z=0-10$ . *ArXiv e-prints*, June 2018.
- P. S. Behroozi and J. Silk. A Simple Technique for Predicting High-redshift Galaxy Evolution. *ApJ*, 799:32, Jan. 2015. doi: 10.1088/0004-637X/799/1/32.
- P. S. Behroozi, C. Conroy, and R. H. Wechsler. A Comprehensive Analysis of Uncertainties Affecting the Stellar Mass-Halo Mass Relation for  $0 < z < 4$ . *ApJ*, 717(1):379–403, July 2010. doi: 10.1088/0004-637X/717/1/379.
- P. S. Behroozi, R. H. Wechsler, and C. Conroy. The Average Star Formation Histories of Galaxies in Dark Matter Halos from  $z = 0-8$ . *ApJ*, 770:57, June 2013. doi: 10.1088/0004-637X/770/1/57.
- N. Benitez, R. Dupke, M. Moles, L. Sodre, J. Cenarro, A. Marin-Franch, K. Taylor, D. Cristobal, A. Fernandez-Soto, C. Mendes de Oliveira, et al. J-PAS: The Javalambre-Physics of the Accelerated Universe Astrophysical Survey. *arXiv e-prints*, art. arXiv:1403.5237, Mar. 2014.
- B. A. Benson, P. A. R. Ade, Z. Ahmed, S. W. Allen, K. Arnold, J. E. Austermann, A. N. Bender, L. E. Bleem, J. E. Carlstrom, C. L. Chang, et al. SPT-3G: a next-generation cosmic microwave background polarization experiment on the South Pole telescope. In *Proc. SPIE*, volume 9153 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 91531P, July 2014. doi: 10.1117/12.2057305.
- F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro. Large-scale structure of the Universe and cosmological perturbation theory. *Phys. Rep.*, 367(1-3):1–248, Sept. 2002. doi: 10.1016/S0370-1573(02)00135-7.
- S. Bhattacharya, K. Heitmann, M. White, Z. Lukić, C. Wagner, and S. Habib. Mass Function Predictions Beyond  $\Lambda$ CDM. *ApJ*, 732:122, May 2011. doi: 10.1088/0004-637X/732/2/122.

- BICEP2 Collaboration, P. A. R. Ade, R. W. Aikin, D. Barkats, S. J. Benton, C. A. Bischoff, J. J. Bock, J. A. Brevik, I. Buder, E. Bullock, et al. Detection of B-Mode Polarization at Degree Angular Scales by BICEP2. *Phys. Rev. Lett.*, 112(24):241101, June 2014. doi: 10.1103/PhysRevLett.112.241101.
- B. Binggeli, A. Sandage, and G. A. Tammann. The luminosity function of galaxies. *ARA&A*, 26:509–560, 1988. doi: 10.1146/annurev.aa.26.090188.002453.
- J. Binney. The physics of dissipational galaxy formation. *ApJ*, 215:483–491, July 1977. doi: 10.1086/155378.
- S. Bird, M. Viel, and M. G. Haehnelt. Massive neutrinos and the non-linear matter power spectrum. *MNRAS*, 420(3):2551–2561, Mar. 2012. doi: 10.1111/j.1365-2966.2011.20222.x.
- Y. Birnboim and A. Dekel. Virial shocks in galactic haloes? *MNRAS*, 345:349–364, Oct. 2003. doi: 10.1046/j.1365-8711.2003.06955.x.
- A. Blanchard and J. Schneider. Gravitational lensing effect on the fluctuations of the cosmic background radiation. *A&A*, 184(1-2):1–6, Oct. 1987.
- D. Blas, J. Lesgourgues, and T. Tram. The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes. *J. Cosmology Astropart. Phys.*, 2011(7):034, July 2011. doi: 10.1088/1475-7516/2011/07/034.
- L. Blot, P. S. Corasaniti, J. M. Alimi, V. Reverdy, and Y. Rasera. Matter power spectrum covariance matrix from the DEUS-PUR  $\Lambda$ CDM simulations: mass resolution and non-Gaussian errors. *MNRAS*, 446(2):1756–1764, Jan. 2015. doi: 10.1093/mnras/stu2190.
- S. Bocquet, A. Saro, K. Dolag, and J. J. Mohr. Halo mass function: baryon impact, fitting formulae, and implications for cluster cosmology. *MNRAS*, 456:2361–2373, Mar. 2016. doi: 10.1093/mnras/stv2657.
- A. D. Bolatto, M. Wolfire, and A. K. Leroy. The CO-to-H<sub>2</sub> Conversion Factor. *ARA&A*, 51(1):207–268, Aug. 2013. doi: 10.1146/annurev-astro-082812-140944.
- J. R. Bond and G. Efstathiou. Cosmic background radiation anisotropies in universes dominated by nonbaryonic dark matter. *ApJ*, 285:L45–L48, Oct. 1984. doi: 10.1086/184362.
- J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser. Excursion Set Mass Functions for Hierarchical Gaussian Fluctuations. *ApJ*, 379:440, Oct. 1991. doi: 10.1086/170520.
- P. Bonifacio, G. Dalton, S. Trager, A. L. Aguerri, E. Carrasco, A. Vallenari, D. C. Abrams, K. Middleton, and F. Sayède. WEAVE an overview and status update. In C. Reylé, J. Richard, L. Cambrésy, M. Deleuil, E. Pécontal, L. Tresse, and I. Vauglin, editors, *SF2A-2016: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, pages 267–270, Dec. 2016.
- G. L. Bryan and M. L. Norman. Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons. *ApJ*, 495:80–99, Mar. 1998. doi: 10.1086/305262.

## BIBLIOGRAPHY

---

- S. Camera, J. Fonseca, R. Maartens, and M. G. Santos. Optimized angular power spectra for spectroscopic galaxy surveys. *MNRAS*, 481(1):1251–1261, Nov. 2018. doi: 10.1093/mnras/sty2284.
- C. L. Carilli and F. Walter. Cool Gas in High-Redshift Galaxies. *ARA&A*, 51(1):105–161, Aug. 2013. doi: 10.1146/annurev-astro-082812-140953.
- J. Carron. On the assumption of Gaussianity for cosmological two-point statistics and parameter dependent covariance matrices. *A&A*, 551:A88, Mar. 2013. doi: 10.1051/0004-6361/201220538.
- A. J. Cenarro, M. Moles, D. Cristóbal-Hornillos, A. Marín-Franch, A. Ederoclite, J. Varela, C. López-Sanjuan, C. Hernández-Monteagudo, R. E. Angulo, H. Vázquez Ramió, et al. J-PLUS: The Javalambre Photometric Local Universe Survey. *A&A*, 622:A176, Feb. 2019. doi: 10.1051/0004-6361/201833036.
- A. Challinor and A. Lewis. Linear power spectrum of observed source number counts. *Phys. Rev. D*, 84(4):043516, Aug. 2011. doi: 10.1103/PhysRevD.84.043516.
- J. Chaves-Montero, C. Hernández-Monteagudo, R. E. Angulo, and J. D. Emberson. Tomographic detection of the kinematic Sunyaev-Zel’dovich effect using angular redshift fluctuations. *arXiv e-prints*, art. arXiv:1911.10690, Nov. 2019.
- M. Chevallier and D. Polarski. Accelerating Universes with Scaling Dark Matter. *International Journal of Modern Physics D*, 10(2):213–223, Jan. 2001. doi: 10.1142/S0218271801000822.
- N. E. Chisari, M. L. A. Richardson, J. Devriendt, Y. Dubois, A. Schneider, A. M. C. Le Brun, R. S. Beckmann, S. Peirani, A. Slyz, and C. Pichon. The impact of baryons on the matter power spectrum from the Horizon-AGN cosmological hydrodynamical simulation. *MNRAS*, 480(3):3962–3977, Nov. 2018. doi: 10.1093/mnras/sty2093.
- C. Cicone, R. Maiolino, S. Gallerani, R. Neri, A. Ferrara, E. Sturm, F. Fiore, E. Piconcelli, and C. Feruglio. Very extended cold gas, star formation and outflows in the halo of a bright QSO at  $z > 6$ . *A&A*, 562:A21, sep 2014. doi: 10.1051/0004-6361/201424980. URL <http://arxiv.org/abs/1409.4418>.
- S. Codis. Nuisance parameters for large galaxy surveys. *arXiv e-prints*, art. arXiv:1612.06873, Dec. 2016.
- S. Cole, P. Norberg, C. M. Baugh, C. S. Frenk, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, et al. The 2dF galaxy redshift survey: near-infrared galaxy luminosity functions. *MNRAS*, 326:255–273, Sept. 2001. doi: 10.1046/j.1365-8711.2001.04591.x.
- C. Conroy, R. H. Wechsler, and A. V. Kravtsov. Modeling Luminosity-dependent Galaxy Clustering through Cosmic Time. *ApJ*, 647:201–214, Aug. 2006. doi: 10.1086/503602.
- A. Cooray and R. Sheth. Halo models of large scale structure. *Phys. Rep.*, 372(1):1–129, Dec. 2002. doi: 10.1016/S0370-1573(02)00276-4.
- T. Costa, D. Sijacki, M. Trenti, and M. Haehnelt. The environment of bright QSOs at  $z \sim 6$ : star-forming galaxies and X-ray emission. *MNRAS*, 439:2146–2174, apr 2014. doi: 10.1093/mnras/stu101.

- M. Costanzi, B. Sartoris, M. Viel, and S. Borgani. Neutrino constraints: what large-scale structure and CMB data are telling us? *J. Cosmology Astropart. Phys.*, 2014(10):081, Oct. 2014. doi: 10.1088/1475-7516/2014/10/081.
- J. Coupon, M. Kilbinger, H. J. McCracken, O. Ilbert, S. Arnouts, Y. Mellier, U. Abbas, S. de la Torre, Y. Goranova, P. Hudelot, et al. Galaxy clustering in the CFHTLS-Wide: the changing relationship between galaxies and haloes since  $z \sim 1.2$ . *A&A*, 542:A5, June 2012. doi: 10.1051/0004-6361/201117625.
- J. Coupon, S. Arnouts, L. van Waerbeke, T. Moutard, O. Ilbert, E. van Uitert, T. Erben, B. Garilli, L. Guzzo, C. Heymans, et al. The galaxy-halo connection from a joint lensing, clustering and abundance analysis in the CFHTLenS/VIPERS field. *MNRAS*, 449:1352–1379, May 2015. doi: 10.1093/mnras/stv276.
- W. I. Cowley, K. I. Caputi, S. Deshmukh, M. L. N. Ashby, G. G. Fazio, O. Le Fèvre, J. P. U. Fynbo, O. Ilbert, H. J. McCracken, B. Milvang-Jensen, et al. The Galaxy-Halo Connection for  $1.5 \lesssim z \lesssim 5$  as Revealed by the Spitzer Matching Survey of the UltraVISTA Ultra-deep Stripes. *ApJ*, 853(1):69, Jan. 2018. doi: 10.3847/1538-4357/aaa41d.
- A. J. Cuesta, M. Vargas-Magaña, F. Beutler, A. S. Bolton, J. R. Brownstein, D. J. Eisenstein, H. Gil-Marín, S. Ho, C. K. McBride, C. Maraston, et al. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the correlation function of LOWZ and CMASS galaxies in Data Release 12. *MNRAS*, 457(2):1770–1785, Apr. 2016. doi: 10.1093/mnras/stw066.
- I. Davidzon, O. Ilbert, C. Laigle, J. Coupon, H. J. McCracken, I. Delvecchio, D. Masters, P. Capak, B. C. Hsieh, O. Le Fèvre, et al. The COSMOS2015 galaxy stellar mass function . Thirteen billion years of stellar mass assembly in ten snapshots. *A&A*, 605: A70, Sept. 2017. doi: 10.1051/0004-6361/201730419.
- I. Davidzon, C. Laigle, P. L. Capak, O. Ilbert, D. C. Masters, S. Hemmati, N. Apostolacos, J. Coupon, S. de la Torre, J. Devriendt, et al. HORIZON-AGN virtual observatory - 2. Template-free estimates of galaxy properties from colours. *MNRAS*, 489(4):4817–4835, Nov. 2019. doi: 10.1093/mnras/stz2486.
- J. J. Davies, R. A. Crain, I. G. McCarthy, B. D. Oppenheimer, J. Schaye, M. Schaller, and S. McAlpine. The gas fractions of dark matter haloes hosting simulated  $\sim L^*$  galaxies are governed by the feedback history of their black holes. *MNRAS*, 485(3):3783–3793, May 2019. doi: 10.1093/mnras/stz635.
- R. D. Davies and A. Wilkinson. Synchrotron Emission from the Galaxy. *arXiv e-prints*, art. astro-ph/9804208, Apr. 1998.
- M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White. The evolution of large-scale structure in a universe dominated by cold dark matter. *ApJ*, 292:371–394, May 1985. doi: 10.1086/163168.
- R. S. de Jong and 4MOST Consortium. 4MOST - The new wide-field, high-multiplex spectroscopic survey facility for ESO’s VISTA telescope. In *IAU General Assembly*, volume 29, page 2255843, Aug. 2015.

## BIBLIOGRAPHY

---

- V. de Lapparent, M. J. Geller, and J. P. Huchra. A Slice of the Universe. *ApJ*, 302:L1, Mar. 1986. doi: 10.1086/184625.
- G. De Lucia, S. Weinmann, B. M. Poggianti, A. Aragón-Salamanca, and D. Zaritsky. The environmental history of group and cluster galaxies in a  $\Lambda$  cold dark matter universe. *MNRAS*, 423:1277–1292, June 2012. doi: 10.1111/j.1365-2966.2012.20983.x.
- A. Dekel and Y. Birnboim. Galaxy bimodality due to cold flows and shock heating. *MNRAS*, 368:2–20, May 2006. doi: 10.1111/j.1365-2966.2006.10145.x.
- A. Dekel, Y. Birnboim, G. Engel, J. Freundlich, T. Goerdt, M. Mumcuoglu, E. Neistein, C. Pichon, R. Teyssier, and E. Zinger. Cold streams in early massive hot haloes as the main mode of galaxy formation. *Nature*, 457:451–454, Jan. 2009. doi: 10.1038/nature07648.
- I. Delvecchio, V. Smolčić, G. Zamorani, C. D. P. Lagos, S. Berta, J. Delhaize, N. Baran, D. M. Alexander, D. J. Rosario, V. Gonzalez-Perez, et al. The VLA-COSMOS 3 GHz Large Project: AGN and host-galaxy properties out to  $z \lesssim 6$ . *A&A*, 602:A3, June 2017. doi: 10.1051/0004-6361/201629367.
- DESI Collaboration, A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, L. E. Allen, C. Allende Prieto, J. Annis, S. Bailey, C. Balland, et al. The DESI Experiment Part I: Science, Targeting, and Survey Design. *arXiv e-prints*, art. arXiv:1611.00036, Oct 2016.
- V. Desjacques, D. Jeong, and F. Schmidt. Large-scale galaxy bias. *Phys. Rep.*, 733:1–193, Feb. 2018. doi: 10.1016/j.physrep.2017.12.002.
- G. Despali, C. Giocoli, R. E. Angulo, G. Tormen, R. K. Sheth, G. Baso, and L. Moscardini. The universality of the virial halo mass function and models for non-universality of other halo definitions. *MNRAS*, 456(3):2486–2504, Mar. 2016. doi: 10.1093/mnras/stv2842.
- E. Di Dio, F. Montanari, R. Durrer, and J. Lesgourgues. Cosmological parameter estimation with large scale structure observations. *J. Cosmology Astropart. Phys.*, 2014(1):042, Jan. 2014. doi: 10.1088/1475-7516/2014/01/042.
- B. Diemer. COLOSSUS: A Python Toolkit for Cosmology, Large-scale Structure, and Dark Matter Halos. *The Astrophysical Journal Supplement Series*, 239(2):35, Dec 2018. doi: 10.3847/1538-4365/aee8c.
- S. Dodelson. *Modern cosmology*. 2003.
- S. Dodelson. *Gravitational Lensing*. 2017.
- S. Dodelson and M. D. Schneider. The effect of covariance estimator error on cosmological parameter constraints. *Phys. Rev. D*, 88(6):063537, Sept. 2013. doi: 10.1103/PhysRevD.88.063537.
- A. D. Dolgov. Neutrinos in cosmology. *Phys. Rep.*, 370(4-5):333–535, Nov. 2002. doi: 10.1016/S0370-1573(02)00139-4.
- O. Doré, J. Bock, M. Ashby, P. Capak, A. Cooray, R. de Putter, T. Eifler, N. Flagey, Y. Gong, S. Habib, et al. Cosmology with the SPHEREX All-Sky Spectral Survey. *arXiv e-prints*, art. arXiv:1412.4872, Dec. 2014.

- M. Douspis, P. G. Castro, C. Caprini, and N. Aghanim. Optimising large galaxy surveys for ISW detection. *A&A*, 485(2):395–401, July 2008. doi: 10.1051/0004-6361/200809499.
- B. T. Draine. Interstellar Dust. In A. McWilliam and M. Rauch, editors, *Origin and Evolution of the Elements*, page 317, Jan. 2004.
- Y. Dubois, C. Pichon, J. Devriendt, J. Silk, M. Haehnelt, T. Kimm, and A. Slyz. Blowing cold flows away: the impact of early AGN activity on the formation of a brightest cluster galaxy progenitor. *MNRAS*, 428:2885–2900, jun 2013. doi: 10.1093/mnras/sts224. URL <http://arxiv.org/abs/1206.5838>.
- Y. Dubois, C. Pichon, C. Welker, D. Le Borgne, J. Devriendt, C. Laigle, S. Codis, D. Pogosyan, S. Arnouts, K. Benabed, et al. Dancing in the dark: galactic properties trace spin swings along the cosmic web. *MNRAS*, 444:1453–1468, Oct. 2014. doi: 10.1093/mnras/stu1227.
- K. Duncan, C. J. Conselice, A. Mortlock, W. G. Hartley, Y. Guo, H. C. Ferguson, R. Davé, Y. Lu, J. Owers, M. L. N. Ashby, et al. The mass evolution of the first galaxies: stellar mass functions and star formation rates at  $4 < z < 7$  in the CANDELS GOODS-South field. *MNRAS*, 444:2960–2984, Nov. 2014. doi: 10.1093/mnras/stu1622.
- F. X. Dupé, A. Rassat, J. L. Starck, and M. J. Fadili. Measuring the integrated Sachs-Wolfe effect. *A&A*, 534:A51, Oct. 2011. doi: 10.1051/0004-6361/201015893.
- A. Durkalec, O. Le Fèvre, A. Pollo, S. de la Torre, P. Cassata, B. Garilli, V. Le Brun, B. C. Lemaux, D. Maccagni, L. Pentericci, et al. Evolution of clustering length, large-scale bias, and host halo mass at  $2 < z < 5$  in the VIMOS Ultra Deep Survey (VUDS)\*. *A&A*, 583:A128, Nov. 2015. doi: 10.1051/0004-6361/201425343.
- A. S. Eddington. On a formula for correcting statistics for the effects of a known error of observation. *MNRAS*, 73:359–360, Mar. 1913. doi: 10.1093/mnras/73.5.359.
- G. Efstathiou, R. S. Ellis, and B. A. Peterson. Analysis of a complete galaxy redshift survey. II - The field-galaxy luminosity function. *MNRAS*, 232:431–461, May 1988. doi: 10.1093/mnras/232.2.431.
- T. Eifler, P. Schneider, and J. Hartlap. Dependence of cosmic shear covariances on cosmology. Impact on parameter estimation. *A&A*, 502(3):721–731, Aug. 2009. doi: 10.1051/0004-6361/200811276.
- A. Einstein. Die Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften (Berlin)*, pages 844–847, Jan. 1915.
- D. J. Eisenstein, W. Hu, and M. Tegmark. Cosmic Complementarity: Joint Parameter Estimation from Cosmic Microwave Background Experiments and Redshift Surveys. *ApJ*, 518(1):2–23, June 1999. doi: 10.1086/307261.
- D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, H.-J. Seo, M. Tegmark, Z. Zheng, et al. Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *ApJ*, 633(2):560–574, Nov. 2005. doi: 10.1086/466512.

## BIBLIOGRAPHY

---

- V. R. Eke, S. Cole, and C. S. Frenk. Cluster evolution as a diagnostic for Omega. *MNRAS*, 282, Sept. 1996. doi: 10.1093/mnras/282.1.263.
- V. R. Eke, C. M. Baugh, S. Cole, C. S. Frenk, and J. F. Navarro. Galaxy groups in the 2dF Galaxy Redshift Survey: the number density of groups. *MNRAS*, 370:1147–1158, Aug. 2006. doi: 10.1111/j.1365-2966.2006.10568.x.
- Euclid Collaboration. Forecasting cosmological constraints for the Euclid and CMB joint analysis. in prep.
- Euclid Collaboration, A. Blanchard, S. Camera, C. Carbone, V. F. Cardone, S. Casas, S. Ilić, M. Kilbinger, T. Kitching, M. Kunz, et al. Euclid preparation: VII. Forecast validation for Euclid cosmological probes. *arXiv e-prints*, art. arXiv:1910.09273, Oct. 2019.
- R. Feldmann and L. Mayer. The Argo simulation - I. Quenching of massive galaxies at high redshift as a result of cosmological starvation. *MNRAS*, 446:1939–1956, Jan. 2015. doi: 10.1093/mnras/stu2207.
- D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. emcee: The MCMC Hammer. *PASP*, 125(925):306, Mar. 2013. doi: 10.1086/670067.
- W. Forman and C. Jones. Hot Gas in Clusters of Galaxies. In W. R. Oegerle, J. Fitchett, and L. Danly, editors, *Clusters of Galaxies*, page 257, Jan. 1990.
- A. Friedmann. Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. *Zeitschrift für Physik*, 21(1):326–332, Dec. 1924. doi: 10.1007/BF01328280.
- Y. Fukuda, T. Hayakawa, E. Ichihara, K. Inoue, K. Ishihara, H. Ishino, Y. Itow, T. Kajita, J. Kameda, S. Kasuga, et al. Measurements of the Solar Neutrino Flux from Super-Kamiokande’s First 300 Days. *Phys. Rev. Lett.*, 81(6):1158–1162, Aug. 1998. doi: 10.1103/PhysRevLett.81.1158.
- J. M. Gabor and R. Davé. Hot gas in massive haloes drives both mass quenching and environment quenching. *MNRAS*, 447:374–391, Feb. 2015. doi: 10.1093/mnras/stu2399.
- T. Giannantonio and W. J. Percival. Using correlations between cosmic microwave background lensing and large-scale structure to measure primordial non-Gaussianity. *MNRAS*, 441:L16–L20, June 2014. doi: 10.1093/mnrasl/slu036.
- T. Giannantonio, A. J. Ross, W. J. Percival, R. Crittenden, D. Bacher, M. Kilbinger, R. Nichol, and J. Weller. Improved primordial non-Gaussianity constraints from measurements of galaxy clustering and the integrated Sachs-Wolfe effect. *Phys. Rev. D*, 89(2):023511, Jan. 2014. doi: 10.1103/PhysRevD.89.023511.
- E. Giusarma, S. Vagnozzi, S. Ho, S. Ferraro, K. Freese, R. Kamen-Rubio, and K.-B. Luk. Scale-dependent galaxy bias, CMB lensing-galaxy cross-correlation, and neutrino masses. *Phys. Rev. D*, 98(12):123526, Dec. 2018. doi: 10.1103/PhysRevD.98.123526.
- M. C. Gonzalez-Garcia, M. Maltoni, J. Salvado, and T. Schwetz. Global fit to three neutrino mixing: critical look at present precision. *Journal of High Energy Physics*, 2012:123, Dec. 2012. doi: 10.1007/JHEP12(2012)123.

- J. Goodman and J. Weare. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5(1):65–80, Jan. 2010. doi: 10.2140/camcos.2010.5.65.
- G. Gozaliasl, A. Finoguenov, H. G. Khosroshahi, B. M. B. Henriques, M. Tanaka, O. Ilbert, S. Wuyts, H. J. McCracken, and F. Montanari. Brightest group galaxies-ii: the relative contribution of bggs to the total baryon content of groups at  $z < 1.3$ . *Monthly Notices of the Royal Astronomical Society*, 475(2):2787–2808, Apr 2018. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/sty003. arXiv: 1801.01492.
- A. Grazian, A. Fontana, P. Santini, J. S. Dunlop, H. C. Ferguson, M. Castellano, R. Amorin, M. L. N. Ashby, G. Barro, P. Behroozi, et al. The galaxy stellar mass function at  $3.5 \leq z \leq 7.5$  in the CANDELS/UDS, GOODS-South, and HUDF fields. *A&A*, 575:A96, Mar. 2015. doi: 10.1051/0004-6361/201424750.
- J. E. Gunn and I. Gott, J. Richard. On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution. *ApJ*, 176:1, Aug. 1972. doi: 10.1086/151605.
- A. H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23(2):347–356, Jan. 1981. doi: 10.1103/PhysRevD.23.347.
- B. Hadzhiyska, B. D. Sherwin, M. Madhavacheril, and S. Ferraro. Improving small-scale CMB lensing reconstruction. *Phys. Rev. D*, 100(2):023547, July 2019. doi: 10.1103/PhysRevD.100.023547.
- A. J. S. Hamilton. Uncorrelated modes of the non-linear power spectrum. *MNRAS*, 312(2):257–284, Feb. 2000. doi: 10.1046/j.1365-8711.2000.03071.x.
- Y. Harikane, M. Ouchi, Y. Ono, S. Saito, P. Behroozi, S. More, K. Shimasaku, J. Toshikawa, Y.-T. Lin, M. Akiyama, et al. GOLDRUSH. II. Clustering of galaxies at  $z \sim 4-6$  revealed with the half-million dropouts over the 100 deg<sup>2</sup> area corresponding to 1 Gpc<sup>3</sup>. *PASJ*, 70:S11, Jan. 2018. doi: 10.1093/pasj/psx097.
- K. Harrington, T. Marriage, A. Ali, J. W. Appel, C. L. Bennett, F. Boone, M. Brewer, M. Chan, D. T. Chuss, F. Colazo, et al. The Cosmology Large Angular Scale Surveyor. In *Proc. SPIE*, volume 9914 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 99141K, July 2016. doi: 10.1117/12.2233125.
- J. Hartlap, P. Simon, and P. Schneider. Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix. *A&A*, 464(1):399–404, Mar. 2007. doi: 10.1051/0004-6361:20066170.
- G. Hasinger, P. Capak, M. Salvato, A. J. Barger, L. L. Cowie, A. Faisst, S. Hemmati, Y. Kakazu, J. Kartaltepe, D. Masters, et al. The DEIMOS 10K Spectroscopic Survey Catalog of the COSMOS Field. *ApJ*, 858:77, May 2018. doi: 10.3847/1538-4357/aabacf.
- S. W. Hawking. The development of irregularities in a single bubble inflationary universe. *Physics Letters B*, 115(4):295–297, Sept. 1982. doi: 10.1016/0370-2693(82)90373-2.
- T. M. Heckman and P. N. Best. The Coevolution of Galaxies and Supermassive Black Holes: Insights from Surveys of the Contemporary Universe. *ARA&A*, 52:589–660, Aug. 2014. doi: 10.1146/annurev-astro-081913-035722.

## BIBLIOGRAPHY

---

- S. W. Henderson, R. Allison, J. Austermann, T. Baildon, N. Battaglia, J. A. Beall, D. Becker, F. De Bernardis, J. R. Bond, E. Calabrese, et al. Advanced ACTPol Cryogenic Detector Arrays and Readout. *Journal of Low Temperature Physics*, 184(3-4): 772–779, Aug. 2016. doi: 10.1007/s10909-016-1575-z.
- C. Hernández-Monteagudo, J. Chaves-Montero, and R. E. Angulo. Angular Redshift Fluctuations: a New Cosmological Observable. *arXiv e-prints*, art. arXiv:1911.12056, Nov. 2019.
- C. Hernández-Monteagudo, J. Chaves-Montero, R. E. Angulo, and G. Ariccò. Tomographic Constraints on Gravity from Angular Redshift Fluctuations in the Late Universe. *arXiv e-prints*, art. arXiv:2005.06568, May 2020.
- W. Hu and B. Jain. Joint galaxy-lensing observables and the dark energy. *Phys. Rev. D*, 70(4):043009, Aug. 2004. doi: 10.1103/PhysRevD.70.043009.
- W. Hu, D. J. Eisenstein, and M. Tegmark. Weighing Neutrinos with Galaxy Surveys. *Phys. Rev. Lett.*, 80(24):5255–5258, June 1998. doi: 10.1103/PhysRevLett.80.5255.
- E. Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15(3):168–173, Mar. 1929. doi: 10.1073/pnas.15.3.168.
- D. Huterer, L. Knox, and R. C. Nichol. The Angular Power Spectrum of Edinburgh/Durham Southern Galaxy Catalogue Galaxies. *ApJ*, 555(2):547–557, July 2001. doi: 10.1086/323328.
- O. Ilbert, M. Salvato, E. Le Floch, H. Aussel, P. Capak, H. J. McCracken, B. Mobasher, J. Kartaltepe, N. Scoville, D. B. Sanders, et al. Galaxy Stellar Mass Assembly Between  $0.2 < z < 2$  from the S-COSMOS Survey. *ApJ*, 709:644–663, Feb. 2010. doi: 10.1088/0004-637X/709/2/644.
- O. Ilbert, H. J. McCracken, O. Le Fèvre, P. Capak, J. Dunlop, A. Karim, M. A. Renzini, K. Caputi, S. Boissier, S. Arnouts, et al. Mass assembly in quiescent and star-forming galaxies since  $z \simeq 4$  from UltraVISTA. *A&A*, 556:A55, Aug. 2013. doi: 10.1051/0004-6361/201321100.
- S. Ishikawa, N. Kashikawa, J. Toshikawa, M. Tanaka, T. Hamana, Y. Niino, K. Ichikawa, and H. Uchiyama. The Galaxy-Halo Connection in High-redshift Universe: Details and Evolution of Stellar-to-halo Mass Ratios of Lyman Break Galaxies on CFHTLS Deep Fields. *ApJ*, 841:8, May 2017. doi: 10.3847/1538-4357/aa6d64.
- Ž. Ivezić, S. M. Kahn, J. A. Tyson, B. Abel, E. Acosta, R. Allsman, D. Alonso, Y. AlSayyad, S. F. Anderson, J. Andrew, et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *ApJ*, 873(2):111, Mar. 2019. doi: 10.3847/1538-4357/ab042c.
- N. Kaiser. On the spatial correlations of Abell clusters. *ApJ*, 284:L9–L12, Sept. 1984. doi: 10.1086/184341.
- N. Kaiser. Clustering in real space and in redshift space. *MNRAS*, 227:1–21, July 1987. doi: 10.1093/mnras/227.1.1.

- R. Keisler, S. Hoover, N. Harrington, J. W. Henning, P. A. R. Ade, K. A. Aird, J. E. Austermann, J. A. Beall, A. N. Bender, B. A. Benson, et al. Measurements of Sub-degree B-mode Polarization in the Cosmic Microwave Background from 100 Square Degrees of SPTpol Data. *ApJ*, 807(2):151, July 2015. doi: 10.1088/0004-637X/807/2/151.
- L. S. Kelvin, S. P. Driver, A. S. G. Robotham, E. N. Taylor, A. W. Graham, M. Alpaslan, I. Baldry, S. P. Bamford, A. E. Bauer, J. Bland-Hawthorn, et al. Galaxy And Mass Assembly (GAMA): stellar mass functions by Hubble type. *MNRAS*, 444:1647–1659, Oct. 2014. doi: 10.1093/mnras/stu1507.
- R. C. Kennicutt and N. J. Evans. Star Formation in the Milky Way and Nearby Galaxies. *ARA&A*, 50:531–608, Sept. 2012. doi: 10.1146/annurev-astro-081811-125610.
- D. Kereš, N. Katz, D. H. Weinberg, and R. Davé. How do galaxies get their gas? *MNRAS*, 363:2–28, Oct. 2005. doi: 10.1111/j.1365-2966.2005.09451.x.
- M. Kilbinger. Cosmology with cosmic shear observations: a review. *Reports on Progress in Physics*, 78(8):086901, July 2015. doi: 10.1088/0034-4885/78/8/086901.
- T. Kitayama and Y. Suto. Formation rate of gravitational structures and the cosmic X-ray background radiation. *MNRAS*, 280:638, May 1996. doi: 10.1093/mnras/280.3.638.
- T. D. Kitching, A. Amara, F. B. Abdalla, B. Joachimi, and A. Refregier. Cosmological systematics beyond nuisance parameters: form-filling functions. *MNRAS*, 399(4):2107–2128, Nov. 2009. doi: 10.1111/j.1365-2966.2009.15408.x.
- A. Knebe, S. R. Knollmann, S. I. Muldrew, F. R. Pearce, M. A. Aragon-Calvo, Y. Ascasibar, P. S. Behroozi, D. Ceverino, S. Colombi, J. Diemand, et al. Haloes gone MAD: The Halo-Finder Comparison Project. *MNRAS*, 415:2293–2318, Aug. 2011. doi: 10.1111/j.1365-2966.2011.18858.x.
- L. Knox. Determination of inflationary observables by cosmic microwave background anisotropy experiments. *Phys. Rev. D*, 52(8):4307–4318, Oct. 1995. doi: 10.1103/PhysRevD.52.4307.
- A. Kogut. Anomalous Microwave Emission. In A. de Oliveira-Costa and M. Tegmark, editors, *Microwave Foregrounds*, volume 181 of *Astronomical Society of the Pacific Conference Series*, page 91, Jan. 1999.
- E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R. Nolte, L. Page, et al. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *ApJS*, 192:18, Feb. 2011. doi: 10.1088/0067-0049/192/2/18.
- E. Krause and T. Eifler. cosmolike - cosmological likelihood analyses for photometric galaxy surveys. *MNRAS*, 470(2):2100–2112, Sept. 2017. doi: 10.1093/mnras/stx1261.
- A. V. Kravtsov, D. Nagai, and A. A. Vikhlinin. Effects of cooling and star formation on the baryon fractions in clusters. *The Astrophysical Journal*, 625(2):588–598, Jun 2005. ISSN 0004-637X. doi: 10.1086/429796.
- F. Lacasa. Covariance of the galaxy angular power spectrum with the halo model. *A&A*, 615:A1, July 2018. doi: 10.1051/0004-6361/201732343.

## BIBLIOGRAPHY

---

- F. Lacasa. Cosmology in the non-linear regime : the small scale miracle. *arXiv e-prints*, art. arXiv:1912.06906, Dec. 2019.
- F. Lacasa and R. Rosenfeld. Combining cluster number counts and galaxy clustering. *J. Cosmology Astropart. Phys.*, 2016(8):005, Aug. 2016. doi: 10.1088/1475-7516/2016/08/005.
- C. Lacey and S. Cole. Merger rates in hierarchical models of galaxy formation. *MNRAS*, 262(3):627–649, June 1993. doi: 10.1093/mnras/262.3.627.
- C. Lacey and S. Cole. Merger Rates in Hierarchical Models of Galaxy Formation - Part Two - Comparison with N-Body Simulations. *MNRAS*, 271:676, Dec. 1994. doi: 10.1093/mnras/271.3.676.
- G. Lagache, J.-L. Puget, and H. Dole. Dusty Infrared Galaxies: Sources of the Cosmic Infrared Background. *ARA&A*, 43(1):727–768, Sept. 2005. doi: 10.1146/annurev.astro.43.072103.150606.
- O. Lahav, P. B. Lilje, J. R. Primack, and M. J. Rees. Dynamical effects of the cosmological constant. *MNRAS*, 251:128–136, July 1991. doi: 10.1093/mnras/251.1.128.
- C. Laigle, H. J. McCracken, O. Ilbert, B. C. Hsieh, I. Davidzon, P. Capak, G. Hasinger, J. D. Silverman, C. Pichon, J. Coupon, et al. The COSMOS2015 Catalog: Exploring the 1 <math>z < 6</math> Universe with Half a Million Galaxies. *ApJS*, 224(2):24, June 2016. doi: 10.3847/0067-0049/224/2/24.
- C. Laigle, I. Davidzon, O. Ilbert, J. Devriendt, D. Kashino, C. Pichon, P. Capak, S. Arnouts, S. de la Torre, Y. Dubois, et al. Horizon-AGN virtual observatory - 1. SED-fitting performance and forecasts for future imaging surveys. *MNRAS*, 486(4): 5104–5123, July 2019. doi: 10.1093/mnras/stz1054.
- D. Larson, J. Dunkley, G. Hinshaw, E. Komatsu, M. R. Nolta, C. L. Bennett, B. Gold, M. Halpern, R. S. Hill, N. Jarosik, et al. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Power Spectra and WMAP-derived Parameters. *ApJS*, 192(2):16, Feb. 2011. doi: 10.1088/0067-0049/192/2/16.
- R. B. Larson. Effects of supernovae on the early evolution of galaxies. *MNRAS*, 169: 229–246, Nov. 1974. doi: 10.1093/mnras/169.2.229.
- R. Laureijs, J. Amiaux, S. Arduini, J. L. Auguères, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet, et al. Euclid Definition Study Report. *arXiv e-prints*, art. arXiv:1110.3193, Oct. 2011.
- O. Le Fèvre, G. Vettolani, B. Garilli, L. Tresse, D. Bottini, V. Le Brun, D. Maccagni, J. P. Picat, R. Scaramella, M. Scodreggio, et al. The VIMOS VLT deep survey. First epoch VVDS-deep survey: 11 564 spectra with  $17.5 \leq IAB \leq 24$ , and the redshift distribution over  $0 \leq z \leq 5$ . *A&A*, 439(3):845–862, Sept. 2005. doi: 10.1051/0004-6361:20041960.
- O. Le Fèvre, L. A. M. Tasca, P. Cassata, B. Garilli, V. Le Brun, D. Maccagni, L. Pentericci, R. Thomas, E. Vanzella, G. Zamorani, et al. The VIMOS Ultra-Deep Survey:  $\sim 10\,000$  galaxies with spectroscopic redshifts to study galaxy assembly at early epochs  $2 < z < 6$ . *A&A*, 576:A79, Apr. 2015. doi: 10.1051/0004-6361/201423829.

- A. Leauthaud, J. Tinker, P. S. Behroozi, M. T. Busha, and R. H. Wechsler. A Theoretical Framework for Combining Techniques that Probe the Link Between Galaxies and Dark Matter. *ApJ*, 738:45, Sept. 2011. doi: 10.1088/0004-637X/738/1/45.
- A. Leauthaud, J. Tinker, K. Bundy, P. S. Behroozi, R. Massey, J. Rhodes, M. R. George, J.-P. Kneib, A. Benson, R. H. Wechsler, et al. New Constraints on the Evolution of the Stellar-to-dark Matter Connection: A Combined Analysis of Galaxy-Galaxy Lensing, Clustering, and Stellar Mass Functions from  $z = 0.2$  to  $z = 1$ . *ApJ*, 744:159, Jan. 2012. doi: 10.1088/0004-637X/744/2/159.
- N. Lee, D. B. Sanders, C. M. Casey, S. Toft, N. Z. Scoville, C.-L. Hung, E. Le Floch, O. Ilbert, H. J. Zahid, H. Aussel, et al. A Turnover in the Galaxy Main Sequence of Star Formation at  $M_* \sim 10^{10} M_\odot$  for Redshifts  $z \lesssim 1.3$ . *ApJ*, 801(2):80, Mar. 2015. doi: 10.1088/0004-637X/801/2/80.
- L. Legrand, H. J. McCracken, I. Davidzon, O. Ilbert, J. Coupon, N. Aghanim, M. Douspis, P. L. Capak, O. Le Fèvre, and B. Milvang-Jensen. The COSMOS-UltraVISTA stellar-to-halo mass relation: new insights on galaxy formation efficiency out to  $z \sim 5$ . *MNRAS*, 486(4):5468–5481, July 2019. doi: 10.1093/mnras/stz1198.
- L. Legrand, C. Hernández-Monteagudo, M. Douspis, N. Aghanim, and R. E. Angulo. High resolution tomography for galaxy spectroscopic surveys with Angular Redshift Fluctuations. *arXiv e-prints*, art. arXiv:2007.14412, July 2020.
- J. Lesgourgues and S. Pastor. Massive neutrinos and cosmology. *Phys. Rep.*, 429(6): 307–379, July 2006. doi: 10.1016/j.physrep.2006.04.001.
- M. Levi and Z. Vlah. Massive neutrinos in nonlinear large scale structure: A consistent perturbation theory. *arXiv e-prints*, art. arXiv:1605.09417, May 2016.
- A. Lewis and S. Bridle. Cosmological parameters from CMB and other data: A Monte Carlo approach. *Phys. Rev. D*, 66(10):103511, Nov. 2002. doi: 10.1103/PhysRevD.66.103511.
- A. Lewis and A. Challinor. Weak gravitational lensing of the CMB. *Phys. Rep.*, 429(1): 1–65, June 2006. doi: 10.1016/j.physrep.2006.03.002.
- S. J. Lilly, O. Le Fèvre, A. Renzini, G. Zamorani, M. Scodeggio, T. Contini, C. M. Carollo, G. Hasinger, J.-P. Kneib, A. Iovino, et al. zCOSMOS: A Large VLT/VIMOS Redshift Survey Covering  $0 < z < 3$  in the COSMOS Field. *ApJS*, 172:70–85, Sept. 2007. doi: 10.1086/516589.
- D. N. Limber. The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field. *ApJ*, 117:134, Jan. 1953. doi: 10.1086/145672.
- A. D. Linde. A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. *Physics Letters B*, 108(6):389–393, Feb. 1982. doi: 10.1016/0370-2693(82)91219-9.
- E. V. Linder. Exploring the Expansion History of the Universe. *Phys. Rev. Lett.*, 90(9): 091301, Mar. 2003. doi: 10.1103/PhysRevLett.90.091301.
- E. V. Linder. Cosmic growth history and expansion history. *Phys. Rev. D*, 72(4):043529, Aug. 2005. doi: 10.1103/PhysRevD.72.043529.

## BIBLIOGRAPHY

---

- G.-C. Liu, N. Sugiyama, A. J. Benson, C. G. Lacey, and A. Nusser. Polarization of the Cosmic Microwave Background from Nonuniform Reionization. *ApJ*, 561(2):504–516, Nov. 2001. doi: 10.1086/323255.
- T. Louis, E. Grace, M. Hasselfield, M. Lungu, L. Maurin, G. E. Addison, P. A. R. Ade, S. Aiola, R. Allison, M. Amiri, et al. The Atacama Cosmology Telescope: two-season ACTPol spectra and parameters. *J. Cosmology Astropart. Phys.*, 2017(6):031, June 2017. doi: 10.1088/1475-7516/2017/06/031.
- M. Loverde and N. Afshordi. Extended Limber approximation. *Phys. Rev. D*, 78(12):123506, Dec. 2008. doi: 10.1103/PhysRevD.78.123506.
- F. Lucchin and S. Matarrese. The Effect of Non-Gaussian Statistics on the Mass Multiplicity of Cosmic Structures. *ApJ*, 330:535, July 1988. doi: 10.1086/166492.
- C.-P. Ma and E. Bertschinger. Cosmological Perturbation Theory in the Synchronous and Conformal Newtonian Gauges. *ApJ*, 455:7, Dec. 1995. doi: 10.1086/176550.
- P. Madau and M. Dickinson. Cosmic Star-Formation History. *ARA&A*, 52:415–486, Aug. 2014. doi: 10.1146/annurev-astro-081811-125615.
- R. Maiolino, S. Gallerani, R. Neri, C. Cicone, A. Ferrara, R. Genzel, D. Lutz, E. Sturm, L. J. Tacconi, F. Walter, et al. Evidence of strong quasar feedback in the early Universe. *MNRAS*, 425(1):L66–L70, sep 2012. ISSN 17453925. doi: 10.1111/j.1745-3933.2012.01303.x. URL <http://arxiv.org/abs/1204.2904><http://mnras1.oxfordjournals.org/cgi/doi/10.1111/j.1745-3933.2012.01303.x>.
- A. Man and S. Belli. Star formation quenching in massive galaxies. *Nature Astronomy*, 2:695–697, Sept. 2018. doi: 10.1038/s41550-018-0558-1.
- C. Marinoni and M. J. Hudson. The Mass-to-Light Function of Virialized Systems and the Relationship between Their Optical and X-Ray Properties. *ApJ*, 569:101–111, Apr. 2002. doi: 10.1086/339319.
- P. Martí, R. Miquel, F. J. Castander, E. Gaztañaga, M. Eriksen, and C. Sánchez. Precise photometric redshifts with a narrow-band filter set: the PAU survey at the William Herschel Telescope. *MNRAS*, 442(1):92–109, July 2014. doi: 10.1093/mnras/stu801.
- J. Martinez-Manso, A. H. Gonzalez, M. L. N. Ashby, S. A. Stanford, M. Brodwin, G. P. Holder, and D. Stern. The Spitzer South Pole Telescope Deep-Field Survey: linking galaxies and haloes at  $z = 1.5$ . *MNRAS*, 446:169–194, Jan. 2015. doi: 10.1093/mnras/stu1998.
- T. Matsumura, Y. Akiba, J. Borrill, Y. Chinone, M. Dobbs, H. Fuke, A. Ghribi, M. Hasegawa, K. Hattori, M. Hattori, et al. Mission Design of LiteBIRD. *Journal of Low Temperature Physics*, 176(5-6):733–740, Sept. 2014. doi: 10.1007/s10909-013-0996-1.
- J. Matthee, J. Schaye, R. A. Crain, M. Schaller, R. Bower, and T. Theuns. The origin of scatter in the stellar mass–halo mass relation of central galaxies in the eagle simulation. *Monthly Notices of the Royal Astronomical Society*, 465(2):2381–2396, Feb 2017. ISSN 0035-8711. doi: 10.1093/mnras/stw2884.

- H. J. McCracken, B. Milvang-Jensen, J. Dunlop, M. Franx, J. P. U. Fynbo, O. Le Fèvre, J. Holt, K. I. Caputi, Y. Goranova, F. Buitrago, et al. UltraVISTA: a new ultra-deep near-infrared survey in COSMOS. *A&A*, 544:A156, Aug. 2012. doi: 10.1051/0004-6361/201219507.
- H. J. McCracken, M. Wolk, S. Colombi, M. Kilbinger, O. Ilbert, S. Peirani, J. Coupon, J. Dunlop, B. Milvang-Jensen, K. Caputi, et al. Probing the galaxy-halo connection in UltraVISTA to  $z \sim 2$ . *MNRAS*, 449(1):901–916, May 2015. doi: 10.1093/mnras/stv305.
- P. McDonald and A. Roy. Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS. *J. Cosmology Astropart. Phys.*, 2009(8):020, Aug. 2009. doi: 10.1088/1475-7516/2009/08/020.
- P. M. Merkel and B. M. Schäfer. Parameter constraints from weak-lensing tomography of galaxy shapes and cosmic microwave background fluctuations. *MNRAS*, 469(3):2760–2770, Aug. 2017. doi: 10.1093/mnras/stx1044.
- M. Moles, N. Benítez, J. A. L. Aguerri, E. J. Alfaro, T. Broadhurst, J. Cabrera-Caño, F. J. Castander, J. Cepa, M. Cerviño, D. Cristóbal-Hornillos, et al. The Alhambra Survey: a Large Area Multimedium-Band Optical and Near-Infrared Photometric Survey. *AJ*, 136(3):1325–1339, Sept. 2008. doi: 10.1088/0004-6256/136/3/1325.
- S. More, F. C. van den Bosch, M. Cacciato, H. J. Mo, X. Yang, and R. Li. Satellite kinematics - II. The halo mass-luminosity relation of central galaxies in SDSS. *MNRAS*, 392:801–816, Jan. 2009. doi: 10.1111/j.1365-2966.2008.14095.x.
- B. P. Moster, R. S. Somerville, C. Maulbetsch, F. C. van den Bosch, A. V. Macciò, T. Naab, and L. Oser. Constraints on the Relationship between Stellar Mass and Halo Mass at Low and High Redshift. *ApJ*, 710:903–923, Feb. 2010. doi: 10.1088/0004-637X/710/2/903.
- B. P. Moster, T. Naab, and S. D. M. White. Galactic star formation and accretion histories from matching galaxies to dark matter haloes. *MNRAS*, 428:3121–3138, Feb. 2013. doi: 10.1093/mnras/sts261.
- B. P. Moster, T. Naab, and S. D. M. White. EMERGE - an empirical model for the formation of galaxies since  $z \sim 10$ . *MNRAS*, 477(2):1822–1852, June 2018. doi: 10.1093/mnras/sty655.
- V. F. Mukhanov and G. V. Chibisov. Quantum fluctuations and a nonsingular universe. *Soviet Journal of Experimental and Theoretical Physics Letters*, 33:532, May 1981.
- S. J. Mutch, D. J. Croton, and G. B. Poole. The simplest model of galaxy formation - I. A formation history model of galaxy stellar mass growth. *MNRAS*, 435:2445–2459, Nov. 2013. doi: 10.1093/mnras/stt1453.
- T. Naab and J. P. Ostriker. Theoretical Challenges in Galaxy Formation. *ARA&A*, 55(1): 59–109, Aug. 2017. doi: 10.1146/annurev-astro-081913-040019.
- J. A. Newman, M. C. Cooper, M. Davis, S. M. Faber, A. L. Coil, P. Guhathakurta, D. C. Koo, A. C. Phillips, C. Conroy, A. A. Dutton, et al. The DEEP2 Galaxy Redshift Survey: Design, Observations, Data Reduction, and Redshifts. *ApJS*, 208(1):5, Sept. 2013. doi: 10.1088/0067-0049/208/1/5.

## BIBLIOGRAPHY

---

- A. J. Nishizawa. The integrated Sachs-Wolfe effect and the Rees-Sciama effect. *Progress of Theoretical and Experimental Physics*, 2014(6):06B110, June 2014. doi: 10.1093/ptep/ptu062.
- K. G. Noeske, B. J. Weiner, S. M. Faber, C. Papovich, D. C. Koo, R. S. Somerville, K. Bundy, C. J. Conselice, J. A. Newman, D. Schiminovich, et al. Star Formation in AEGIS Field Galaxies since  $z=1.1$ : The Dominance of Gradually Declining Star Formation, and the Main Sequence of Star-forming Galaxies. *ApJ*, 660(1):L43–L46, May 2007. doi: 10.1086/517926.
- T. Okamoto and W. Hu. Cosmic microwave background lensing reconstruction on the full sky. *Phys. Rev. D*, 67(8):083002, Apr. 2003. doi: 10.1103/PhysRevD.67.083002.
- Y. Omori, R. Chown, G. Simard, K. T. Story, K. Aylor, E. J. Baxter, B. A. Benson, L. E. Bleem, J. E. Carlstrom, C. L. Chang, et al. A 2500 deg<sup>2</sup> CMB Lensing Map from Combined South Pole Telescope and Planck Data. *ApJ*, 849(2):124, Nov. 2017. doi: 10.3847/1538-4357/aa8d1d.
- N. Padmanabhan, D. J. Schlegel, U. Seljak, A. Makarov, N. A. Bahcall, M. R. Blanton, J. Brinkmann, D. J. Eisenstein, D. P. Finkbeiner, J. E. Gunn, et al. The clustering of luminous red galaxies in the Sloan Digital Sky Survey imaging data. *MNRAS*, 378(3): 852–872, July 2007. doi: 10.1111/j.1365-2966.2007.11593.x.
- S. Passaglia, A. Manzotti, and S. Dodelson. Cross-correlating 2D and 3D galaxy surveys. *Phys. Rev. D*, 95(12):123508, June 2017. doi: 10.1103/PhysRevD.95.123508.
- J. A. Peacock. *Cosmological Physics*. Cambridge University Press, 1999.
- R. Pearson and O. Zahn. Cosmology from cross correlation of CMB lensing and galaxy surveys. *Phys. Rev. D*, 89(4):043516, Feb. 2014. doi: 10.1103/PhysRevD.89.043516.
- P. J. E. Peebles. *The large-scale structure of the universe*. 1980.
- A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, 142:419–421, July 1965. doi: 10.1086/148307.
- W. J. Percival, B. A. Reid, D. J. Eisenstein, N. A. Bahcall, T. Budavari, J. A. Frieman, M. Fukugita, J. E. Gunn, Ž. Ivezić, G. R. Knapp, et al. Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample. *MNRAS*, 401(4):2148–2168, Feb. 2010. doi: 10.1111/j.1365-2966.2009.15812.x.
- P. G. Pérez-González, A. Cava, G. Barro, V. Villar, N. Cardiel, I. Ferreras, J. M. Rodríguez-Espinosa, A. Alonso-Herrero, M. Balcells, J. Cenarro, et al. SHARDS: An Optical Spectro-photometric Survey of Distant Galaxies. *ApJ*, 762(1):46, Jan. 2013. doi: 10.1088/0004-637X/762/1/46.
- Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, et al. Planck 2015 results. XIII. Cosmological parameters. *A&A*, 594:A13, Sept. 2016a. doi: 10.1051/0004-6361/201525830.

- Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, et al. Planck 2015 results. XXIV. Cosmology from Sunyaev-Zeldovich cluster counts. *A&A*, 594:A24, Sept. 2016b. doi: 10.1051/0004-6361/201525833.
- Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, et al. Planck 2018 results. VI. Cosmological parameters. *arXiv e-prints*, art. arXiv:1807.06209, July 2018a.
- Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, et al. Planck 2018 results. VIII. Gravitational lensing. *arXiv e-prints*, art. arXiv:1807.06210, July 2018b.
- Planck Collaboration, Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, et al. Planck 2018 results. I. Overview and the cosmological legacy of Planck. *arXiv e-prints*, art. arXiv:1807.06205, July 2018c.
- Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, et al. Planck 2018 results. V. CMB power spectra and likelihoods. *arXiv e-prints*, art. arXiv:1907.12875, July 2019.
- Polarbear Collaboration, P. A. R. Ade, Y. Akiba, A. E. Anthony, K. Arnold, M. Atlas, D. Barron, D. Boettger, J. Borrill, S. Chapman, et al. A Measurement of the Cosmic Microwave Background B-mode Polarization Power Spectrum at Sub-degree Scales with POLARBEAR. *ApJ*, 794(2):171, Oct. 2014. doi: 10.1088/0004-637X/794/2/171.
- D. Potter, J. Stadel, and R. Teyssier. PKDGRAV3: beyond trillion particle cosmological simulations for the next era of galaxy surveys. *Computational Astrophysics and Cosmology*, 4(1):2, May 2017. doi: 10.1186/s40668-017-0021-1.
- L. C. Powell, A. Slyz, and J. Devriendt. The impact of supernova-driven winds on stream-fed protogalaxies. *MNRAS*, 414(4):3671–3689, July 2011. doi: 10.1111/j.1365-2966.2011.18668.x.
- L. Pozzetti, C. M. Hirata, J. E. Geach, A. Cimatti, C. Baugh, O. Cucciati, A. Merson, P. Norberg, and D. Shi. Modelling the number density of H $\alpha$  emitters for future spectroscopic near-IR space missions. *A&A*, 590:A3, May 2016. doi: 10.1051/0004-6361/201527081.
- W. H. Press and P. Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *ApJ*, 187:425–438, Feb. 1974. doi: 10.1086/152650.
- R. M. Reddick, R. H. Wechsler, J. L. Tinker, and P. S. Behroozi. The Connection between Galaxies and Dark Matter Structures in the Local Universe. *ApJ*, 771(1):30, July 2013. doi: 10.1088/0004-637X/771/1/30.
- M. J. Rees and J. P. Ostriker. Cooling, dynamics and fragmentation of massive gas clouds: clues to the masses and radii of galaxies and clusters. *MNRAS*, 179:541–559, June 1977. doi: 10.1093/mnras/179.4.541.

## BIBLIOGRAPHY

---

- A. Rodríguez-Puebla, P. Behroozi, J. Primack, A. Klypin, C. Lee, and D. Hellinger. Halo and subhalo demographics with Planck cosmological parameters: Bolshoi-Planck and MultiDark-Planck simulations. *MNRAS*, 462:893–916, Oct. 2016. doi: 10.1093/mnras/stw1705.
- A. Rodríguez-Puebla, J. R. Primack, V. Avila-Reese, and S. M. Faber. Constraining the galaxy-halo connection over the last 13.3 Gyr: star formation histories, galaxy mergers and structural properties. *MNRAS*, 470:651–687, Sept. 2017. doi: 10.1093/mnras/stx1172.
- D. H. Rudd, A. R. Zentner, and A. V. Kravtsov. Effects of Baryons and Dissipation on the Matter Power Spectrum. *ApJ*, 672(1):19–32, Jan. 2008. doi: 10.1086/523836.
- R. K. Sachs and A. M. Wolfe. Perturbations of a Cosmological Model and Angular Variations of the Microwave Background. *ApJ*, 147:73, Jan. 1967. doi: 10.1086/148982.
- P. Salomé, F. Combes, A. C. Edge, C. Crawford, M. Erlund, A. C. Fabian, N. A. Hatch, R. M. Johnstone, J. S. Sanders, and R. J. Wilman. Cold molecular gas in the Perseus cluster core. Association with X-ray cavity, H $\alpha$  filaments and cooling flow. *A&A*, 454(2):437–445, Aug. 2006. doi: 10.1051/0004-6361:20054745.
- L. Salvati, M. Douspis, and N. Aghanim. Constraints from thermal Sunyaev-Zel’dovich cluster counts and power spectrum combined with CMB. *A&A*, 614:A13, June 2018. doi: 10.1051/0004-6361/201731990.
- A. Sandage, G. A. Tammann, and A. Yahil. The velocity field of bright nearby galaxies. I - The variation of mean absolute magnitude with redshift for galaxies in a magnitude-limited sample. *ApJ*, 232:352–364, Sept. 1979. doi: 10.1086/157295.
- E. Schaan and S. Ferraro. Foreground-Immune Cosmic Microwave Background Lensing with Shear-Only Reconstruction. *Phys. Rev. Lett.*, 122(18):181301, May 2019. doi: 10.1103/PhysRevLett.122.181301.
- E. Schaan, E. Krause, T. Eifler, O. Doré, H. Miyatake, J. Rhodes, and D. N. Spergel. Looking through the same lens: Shear calibration for LSST, Euclid, and WFIRST with stage 4 CMB lensing. *Phys. Rev. D*, 95(12):123512, June 2017. doi: 10.1103/PhysRevD.95.123512.
- J. Schaye, R. A. Crain, R. G. Bower, M. Furlong, M. Schaller, T. Theuns, C. Dalla Vecchia, C. S. Frenk, I. G. McCarthy, J. C. Helly, et al. The EAGLE project: simulating the evolution and assembly of galaxies and their environments. *MNRAS*, 446:521–554, Jan. 2015. doi: 10.1093/mnras/stu2058.
- P. Schechter. An analytic expression for the luminosity function for galaxies. *ApJ*, 203:297–306, Jan. 1976. doi: 10.1086/154079.
- M. Schmidt. Space Distribution and Luminosity Functions of Quasi-Stellar Radio Sources. *ApJ*, 151:393, Feb. 1968. doi: 10.1086/149446.
- A. Schneider and R. Teyssier. A new method to quantify the effects of baryons on the matter power spectrum. *J. Cosmology Astropart. Phys.*, 2015(12):049, Dec. 2015. doi: 10.1088/1475-7516/2015/12/049.

- A. Schneider, R. Teyssier, D. Potter, J. Stadel, J. Onions, D. S. Reed, R. E. Smith, V. Springel, F. R. Pearce, and R. Scoccimarro. Matter power spectrum and the challenge of percent accuracy. *J. Cosmology Astropart. Phys.*, 2016(4):047, Apr. 2016. doi: 10.1088/1475-7516/2016/04/047.
- A. Schneider, N. Stoira, A. Refregier, A. J. Weiss, M. Knabenhans, J. Stadel, and R. Teyssier. Baryonic effects for weak lensing. Part I. Power spectrum and covariance matrix. *J. Cosmology Astropart. Phys.*, 2020(4):019, Apr. 2020. doi: 10.1088/1475-7516/2020/04/019.
- N. Scoville, H. Aussel, A. Benson, A. Blain, D. Calzetti, P. Capak, R. S. Ellis, A. El-Zant, A. Finoguenov, M. Giavalisco, et al. Large Structures and Galaxy Evolution in COSMOS at  $z \lesssim 1.1$ . *ApJS*, 172(1):150–181, Sept. 2007a. doi: 10.1086/516751.
- N. Scoville, H. Aussel, M. Brusa, P. Capak, C. M. Carollo, M. Elvis, M. Giavalisco, L. Guzzo, G. Hasinger, C. Impey, et al. The Cosmic Evolution Survey (COSMOS): Overview. *ApJS*, 172:1–8, Sept. 2007b. doi: 10.1086/516585.
- N. Scoville, K. Sheth, H. Aussel, P. Vanden Bout, P. Capak, A. Bongiorno, C. M. Casey, L. Murchikova, J. Koda, J. Álvarez-Márquez, et al. ISM Masses and the Star formation Law at  $Z = 1$  to 6: ALMA Observations of Dust Continuum in 145 Galaxies in the COSMOS Survey Field. *ApJ*, 820(2):83, Apr. 2016. doi: 10.3847/0004-637X/820/2/83.
- N. Scoville, N. Lee, P. Vanden Bout, T. Diaz-Santos, D. Sanders, B. Darvish, A. Bongiorno, C. M. Casey, L. Murchikova, J. Koda, et al. Evolution of Interstellar Medium, Star Formation, and Accretion at High Redshift. *ApJ*, 837(2):150, Mar. 2017. doi: 10.3847/1538-4357/aa61a0.
- U. Seljak and M. Zaldarriaga. A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies. *ApJ*, 469:437, Oct. 1996. doi: 10.1086/177793.
- E. Sellentin and A. F. Heavens. On the insufficiency of arbitrarily precise covariance matrices: non-Gaussian weak-lensing likelihoods. *MNRAS*, 473(2):2355–2363, Jan. 2018. doi: 10.1093/mnras/stx2491.
- E. Sellentin, M. Quartin, and L. Amendola. Breaking the spell of Gaussianity: forecasting with higher order Fisher matrices. *MNRAS*, 441(2):1831–1840, June 2014. doi: 10.1093/mnras/stu689.
- R. K. Sheth and G. Tormen. Large-scale bias and the peak background split. *MNRAS*, 308(1):119–126, Sept. 1999. doi: 10.1046/j.1365-8711.1999.02692.x.
- J. M. Shull, B. D. Smith, and C. W. Danforth. The Baryon Census in a Multiphase Intergalactic Medium: 30% of the Baryons May Still be Missing. *ApJ*, 759(1):23, Nov. 2012. doi: 10.1088/0004-637X/759/1/23.
- V. F. Shvartsman. Density of relict particles with zero rest mass in the universe. *Soviet Journal of Experimental and Theoretical Physics Letters*, 9:184–186, Jan. 1969.
- J. L. Sievers, R. A. Hlozek, M. R. Nolta, V. Acquaviva, G. E. Addison, P. A. R. Ade, P. Aguirre, M. Amiri, J. W. Appel, L. F. Barrientos, et al. The Atacama Cosmology Telescope: cosmological parameters from three seasons of data. *J. Cosmology Astropart. Phys.*, 2013(10):060, Oct. 2013. doi: 10.1088/1475-7516/2013/10/060.

## BIBLIOGRAPHY

---

- J. Silk. On the fragmentation of cosmic gas clouds. I. The formation of galaxies and the first generation of stars. *ApJ*, 211:638–648, Feb. 1977. doi: 10.1086/154972.
- J. Silk and G. A. Mamon. The current status of galaxy formation. *Research in Astronomy and Astrophysics*, 12:917–946, Aug. 2012. doi: 10.1088/1674-4527/12/8/004.
- J. Silk and M. J. Rees. Quasars and galaxy formation. *A&A*, 331:L1–L4, Mar. 1998.
- J. Silk, A. Di Cintio, and I. Dvorkin. Galaxy Formation. In *Proceedings of the International School of Physics 'Enrico Fermi' Course 186 'New Horizons for Observational Cosmology' Vol. 186*, volume 186, pages 137–187, Dec. 2014. doi: 10.3254/978-1-61499-476-3-137.
- G. F. Smoot, C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, N. W. Boggess, E. S. Cheng, G. de Amici, S. Gulbis, M. G. Hauser, et al. Structure in the COBE Differential Microwave Radiometer First-Year Maps. *ApJ*, 396:L1, Sept. 1992. doi: 10.1086/186504.
- J. S. Speagle, C. L. Steinhardt, P. L. Capak, and J. D. Silverman. A Highly Consistent Framework for the Evolution of the Star-Forming “Main Sequence” from  $z \sim 0$ -6. *ApJS*, 214(2):15, Oct. 2014. doi: 10.1088/0067-0049/214/2/15.
- V. Springel. The cosmological simulation code GADGET-2. *MNRAS*, 364(4):1105–1134, Dec. 2005. doi: 10.1111/j.1365-2966.2005.09655.x.
- G. Steigman, D. N. Schramm, and J. E. Gunn. Cosmological limits to the number of massive leptons. *Physics Letters B*, 66(2):202–204, Jan. 1977. doi: 10.1016/0370-2693(77)90176-9.
- C. L. Steinhardt, J. S. Speagle, P. Capak, J. D. Silverman, M. Carollo, J. Dunlop, Y. Hashimoto, B.-C. Hsieh, O. Ilbert, O. Le Fevre, et al. Star Formation at  $4 < z < 6$  from the Spitzer Large Area Survey with Hyper-Suprime-Cam (SPLASH). *ApJ*, 791:L25, Aug. 2014. doi: 10.1088/2041-8205/791/2/L25.
- K. T. Story, C. L. Reichardt, Z. Hou, R. Keisler, K. A. Aird, B. A. Benson, L. E. Bleem, J. E. Carlstrom, C. L. Chang, H. M. Cho, et al. A Measurement of the Cosmic Microwave Background Damping Tail from the 2500-Square-Degree SPT-SZ Survey. *ApJ*, 779(1):86, Dec. 2013. doi: 10.1088/0004-637X/779/1/86.
- M. A. Strauss and J. A. Willick. The density and peculiar velocity fields of nearby galaxies. *Phys. Rep.*, 261:271–431, Jan. 1995. doi: 10.1016/0370-1573(95)00013-7.
- R. A. Sunyaev and I. B. Zeldovich. Microwave background radiation as a probe of the contemporary structure and history of the universe. *ARA&A*, 18:537–560, Jan. 1980. doi: 10.1146/annurev.aa.18.090180.002541.
- R. A. Sunyaev and Y. B. Zeldovich. The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies. *Comments on Astrophysics and Space Physics*, 4:173, Nov. 1972.
- A. Suzuki, P. Ade, Y. Akiba, C. Aleman, K. Arnold, C. Baccigalupi, B. Barch, D. Barron, A. Bender, D. Boettger, et al. The Polarbear-2 and the Simons Array Experiments. *Journal of Low Temperature Physics*, 184(3-4):805–810, Aug. 2016. doi: 10.1007/s10909-015-1425-4.

- N. Suzuki, D. Rubin, C. Lidman, G. Aldering, R. Amanullah, K. Barbary, L. F. Barrientos, J. Botyanszki, M. Brodwin, N. Connolly, et al. The Hubble Space Telescope Cluster Supernova Survey. V. Improving the Dark-energy Constraints above  $z > 1$  and Building an Early-type-hosted Supernova Sample. *ApJ*, 746(1):85, Feb. 2012. doi: 10.1088/0004-637X/746/1/85.
- L. J. Tacconi, R. Genzel, R. Neri, P. Cox, M. C. Cooper, K. Shapiro, A. Bolatto, N. Bouché, F. Bournaud, A. Burkert, et al. High molecular gas fractions in normal massive star-forming galaxies in the young Universe. *Nature*, 463(7282):781–784, Feb. 2010. doi: 10.1038/nature08773.
- L. J. Tacconi, R. Genzel, A. Saintonge, F. Combes, S. García-Burillo, R. Neri, A. Bolatto, T. Contini, N. M. Förster Schreiber, S. Lilly, et al. PHIBSS: Unified Scaling Relations of Gas Depletion Time and Molecular Gas Fractions. *ApJ*, 853(2):179, Feb. 2018. doi: 10.3847/1538-4357/aaa4b4.
- R. Takahashi, M. Sato, T. Nishimichi, A. Taruya, and M. Oguri. Revising the Halofit Model for the Nonlinear Matter Power Spectrum. *ApJ*, 761(2):152, Dec. 2012. doi: 10.1088/0004-637X/761/2/152.
- A. Taruya, F. Bernardeau, T. Nishimichi, and S. Codis. Direct and fast calculation of regularized cosmological power spectrum at two-loop order. *Phys. Rev. D*, 86(10):103528, Nov. 2012. doi: 10.1103/PhysRevD.86.103528.
- J. A. Tauber, N. Mandolesi, J. L. Puget, T. Banos, M. Bersanelli, F. R. Bouchet, R. C. Butler, J. Charra, G. Crone, J. Dodsworth, et al. Planck pre-launch status: The Planck mission. *A&A*, 520:A1, Sept. 2010. doi: 10.1051/0004-6361/200912983.
- A. Taylor, B. Joachimi, and T. Kitching. Putting the precision in precision cosmology: How accurate should your data covariance matrix be? *MNRAS*, 432(3):1928–1946, July 2013. doi: 10.1093/mnras/stt270.
- M. Tegmark. Measuring Cosmological Parameters with Galaxy Surveys. *Phys. Rev. Lett.*, 79(20):3806–3809, Nov. 1997. doi: 10.1103/PhysRevLett.79.3806.
- M. Tegmark, A. N. Taylor, and A. F. Heavens. Karhunen-Loève Eigenvalue Problems in Cosmology: How Should We Tackle Large Data Sets? *ApJ*, 480(1):22–35, May 1997. doi: 10.1086/303939.
- R. Teyssier. Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES. *A&A*, 385:337–364, Apr. 2002. doi: 10.1051/0004-6361:20011817.
- R. Teyssier, S. Pires, S. Prunet, D. Aubert, C. Pichon, A. Amara, K. Benabed, S. Colombi, A. Refregier, and J. L. Starck. Full-sky weak-lensing simulation with 70 billion particles. *A&A*, 497(2):335–341, Apr. 2009. doi: 10.1051/0004-6361/200810657.
- The Simons Observatory Collaboration, P. Ade, J. Aguirre, Z. Ahmed, S. Aiola, A. Ali, D. Alonso, M. A. Alvarez, K. Arnold, P. Ashton, et al. The Simons Observatory: science goals and forecasts. *J. Cosmology Astropart. Phys.*, 2:056, Feb. 2019. doi: 10.1088/1475-7516/2019/02/056.

## BIBLIOGRAPHY

---

- J. Tinker, A. V. Kravtsov, A. Klypin, K. Abazajian, M. Warren, G. Yepes, S. Gottlöber, and D. E. Holz. Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality. *ApJ*, 688(2):709–728, Dec. 2008. doi: 10.1086/591439.
- L. Toffolatti, G. De Zotti, F. Argüeso, and C. Burigana. Extragalactic Radio Sources and CMB Anisotropies. In A. de Oliveira-Costa and M. Tegmark, editors, *Microwave Foregrounds*, volume 181 of *Astronomical Society of the Pacific Conference Series*, page 153, Jan. 1999.
- W. H. Tucker and R. J. Gould. Radiation from a Low-Density Plasma at  $10^6$  ° -  $10^8$  ° K. *ApJ*, 144:244, Apr. 1966. doi: 10.1086/148601.
- R. B. Tully and J. R. Fisher. Reprint of 1977A&A....54..661T. A new method of determining distance to galaxies. *A&A*, 500:105–117, Feb. 1977.
- A. Vale and J. P. Ostriker. Linking halo mass to galaxy luminosity. *MNRAS*, 353:189–200, Sept. 2004. doi: 10.1111/j.1365-2966.2004.08059.x.
- A. Vallinotto. The Synergy between the Dark Energy Survey and the South Pole Telescope. *ApJ*, 778(2):108, Dec. 2013. doi: 10.1088/0004-637X/778/2/108.
- L. Verde. Statistical Methods in Cosmology. 800:147–177, 2010. doi: 10.1007/978-3-642-10598-2\_4.
- N. Vittorio and J. Silk. Fine-scale anisotropy of the cosmic microwave background in a universe dominated by cold dark matter. *ApJ*, 285:L39–L43, Oct. 1984. doi: 10.1086/184361.
- W. A. Watson, I. T. Iliev, A. D’Aloisio, A. Knebe, P. R. Shapiro, and G. Yepes. The halo mass function through the cosmic ages. *MNRAS*, 433:1230–1245, Aug. 2013. doi: 10.1093/mnras/stt791.
- M. White. The mass of a halo. *A&A*, 367:27–32, Feb. 2001. doi: 10.1051/0004-6361:20000357.
- L. Wolz, M. Kilbinger, J. Weller, and T. Giannantonio. On the validity of cosmological Fisher matrix forecasts. *J. Cosmology Astropart. Phys.*, 2012(9):009, Sept. 2012. doi: 10.1088/1475-7516/2012/09/009.
- A. H. Wright, S. P. Driver, and A. S. G. Robotham. GAMA/G10-COSMOS/3D-HST: Evolution of the galaxy stellar mass function over 12.5 Gyr. *MNRAS*, 480:3491–3502, Nov. 2018. doi: 10.1093/mnras/sty2136.
- X. Yang, H. J. Mo, and F. C. van den Bosch. Constraining galaxy formation and cosmology with the conditional luminosity function of galaxies. *MNRAS*, 339:1057–1080, Mar. 2003. doi: 10.1046/j.1365-8711.2003.06254.x.
- X. Yang, H. J. Mo, F. C. van den Bosch, Y. Zhang, and J. Han. Evolution of the Galaxy-Dark Matter Connection and the Assembly of Galaxies in Dark Matter Halos. *ApJ*, 752:41, June 2012. doi: 10.1088/0004-637X/752/1/41.
- Z. Zheng, A. L. Coil, and I. Zehavi. Galaxy Evolution from Halo Occupation Distribution Modeling of DEEP2 and SDSS Galaxy Clustering. *ApJ*, 667:760–779, Oct. 2007. doi: 10.1086/521074.



**Titre:** Grands relevés: de l'évolution des galaxies aux sondes cosmologiques

**Mots clés:** Galaxies: evolution - halos; Cosmologie: grandes structures de l'Univers - paramètres cosmologiques

**Résumé:** Les grands relevés de galaxies sont des fenêtres ouvertes sur notre Univers: ils nous offrent de précieuses informations sur son contenu et sur son évolution. D'une part les relevés profonds explorent la formation et l'évolution des galaxies. D'autre part, les relevés à grand champ cartographient la distribution de la matière dans le but de comprendre la nature de l'énergie noire et de la matière noire. Au cours de cette thèse, j'explore les capacités offertes par ces relevés afin de répondre aux questions suivantes:

1. Quels sont les principaux moteurs de l'évolutions des galaxies ?
2. Quelles progrès dans notre connaissance de l'Univers seront apportés par les futurs relevés de galaxies ?

Je commence par déterminer la relation entre la masse stellaire et la masse des halos de matière noire des galaxies en utilisant des mesures précises de la fonction de masse stellaire dans le champ COSMOS. Grâce à l'exhaustivité du relevé COSMOS entre  $z \sim 0.2$  et  $z \sim 5$ , j'obtiens pour la première fois cette relation sur une aussi grande gamme de redshifts à partir d'un seul relevé. Je constate que la masse de halo caractéristique, définie comme maximisant le rapport entre la masse stellaire et la masse du halo, augmente entre  $z = 0$  et  $z = 2.3$  et reste stable jusqu'à  $z = 4$ . Cette augmentation de la masse de halo caractéristique met en lumière le rôle des flux de gaz froid comme moteurs de la formation des galaxies à grand redshift. Afin d'approfondir ce sujet, je combine des observations de la teneur en gaz moléculaire froid des galaxies jusqu'à  $z = 4$ , avec la relation entre masse stellaire et masse du halo de matière noire. Je constate que l'évolution de la fraction de masse du gaz froid est en accord avec l'hypothèse

selon laquelle les apports de gaz froid sont responsables de la plus grande efficacité de formation des galaxies à grand redshift dans les halos massifs.

Ensuite, dans le but de maximiser les contraintes cosmologique qui seront apportées par les prochains grands relevés spectroscopiques, je montre qu'une nouvelle observable, les fluctuations angulaires de redshift (ARF), apportent des informations complémentaires par rapport au traditionnel "angular galaxy clustering". Grâce à leurs sensibilités particulières au champ de vitesse de la matière, je montre que les dégénérescences entre les paramètres cosmologiques et de biais des galaxies sont différentes lorsqu'elles se basent sur les ARF ou sur le "angular galaxy clustering". Dès lors, la combinaison des deux observables permet de lever des dégénérescences et d'améliorer les contraintes, d'un facteur au moins deux, sur la plupart des paramètres des modèles  $\Lambda$ CDM et  $w$ CDM.

Finalement, en tant que membre de la collaboration Euclid, j'ai exploré le potentiel cosmologique de ce futur relevé de galaxies. Ce relevé nous permettra de mesurer très précisément la distribution de la matière sur tout le ciel extragalactique. Dans le but d'exploiter entièrement tout son potentiel, il est crucial de le combiner avec les futurs relevés du CMB. J'utilise le formalisme de Fisher afin de prédire l'intérêt d'une analyse combinée des sondes CMB avec les sondes Euclid. Je teste à la fois le modèle  $\Lambda$ CDM standard et ses extensions, et montre que le CMB améliorera les contraintes d'un facteur au moins deux sur la plupart des paramètres cosmologiques, et notamment sur les modèles d'énergie noire alternatifs, qui font partie des intérêts majeurs pour Euclid.

**Title:** Large surveys: from galaxy evolution to cosmological probes

**Keywords:** Galaxies: evolution - haloes; Cosmology: large-scale structure of Universe - cosmological parameters

**Abstract:** Large galaxy surveys are like open windows on our Universe: they provide precious insights on its components and on its evolution. On the one hand, pencil surveys go deep into the cosmos to explore the formation and evolution of galaxies. On the other hand, wide surveys are mapping the distribution of matter on large scales to understand the nature of dark energy and dark matter. During my thesis, I explore the capabilities of these large surveys to address the following questions:

1. What are the main drivers of galaxy evolution?
2. What improvements on our knowledge of the Universe will be brought by upcoming wide galaxy surveys?

Using precise galaxy stellar-mass function measurements in the COSMOS field, I first determine the stellar-to-halo mass relation through a parametric abundance matching technique. Thanks to the completeness of the COSMOS survey from  $z \sim 0.2$  to  $z \sim 5$ , I obtain for the first time this relation over such a large redshift range from a single coherent sample. I find that the ratio of stellar-to-halo mass content peaks at a characteristic halo mass which increases up to  $z = 2.3$  and remains flat up to  $z = 4$ . This steady increase of the characteristic halo mass questions the role of cold gas inflows as drivers of galaxy formation at high redshift. To address this question, I link observations of the cold molecular gas content in galaxies up to  $z = 4$  to the evolution of the dark matter halo mass. I find that the joint evolution of cold gas mass fraction and halo mass is in agreement with the hypothesis of

cold gas inflows being responsible of efficient galaxy formation at high redshift.

With the scope of maximising the cosmological power of next generation spectroscopic surveys, I show that a novel cosmological observable, the angular redshift fluctuations (ARF) will provide complementary cosmological information in addition to the standard angular galaxy clustering. Due to its distinct sensitivity to the peculiar velocity field, I find that the cosmological and galaxy bias parameters express different degeneracies when inferred from ARF or from angular galaxy clustering. As such, combining both observables breaks these degeneracies and greatly decreases the marginalised uncertainties, by a factor of at least two on most parameters in the  $\Lambda$ CDM and  $w$ CDM models.

As part of the Euclid collaboration, I then investigate the cosmological power of the upcoming Euclid survey, which will offer us an exquisite measurement of the matter distributions on the full extra-galactic sky. In order to fully exploit all the potential of the Euclid survey it is crucial to combine it with upcoming CMB surveys. I use the Fisher formalism to forecast the benefits of performing a joint analysis of CMB probes with Euclid main probes (weak lensing and galaxy clustering). I test both the standard cosmological model,  $\Lambda$ CDM, and its extensions, and show that CMB will improve the constraints by a factor two on most cosmological parameters, and most notably on dark energy modified models which are of key interest for Euclid.