Università degli Studi di Genova

Scuola di dottorato in Fisica e Nanoscienze

Improving the *Euclid* performance: from spectroscopic simulations to the 6×2pt statistics

PhD thesis

December 17, 2021

PhD Candidate

Luca Paganin

Tutors

Stefano Davini
Istituto Nazionale di Fisica Nucleare

Silvano Tosi
Università degli Studi di Genova

External advisors

Carmelita Carbone
Istituto Nazionale di Astrofisica

Stefano Camera
Università degli Studi di Torino

Università degli Studi di Genova

Ciclo XXXIV – 2018-2021

Contents

Introduction

Cosmology is the field of Physics which studies the Universe as a whole. Our ancestors were already asking themselves profound questions about the cosmos thousands of years ago. However, the birth of Modern Cosmology is quite recent, dating back to the second decade of the twentieth century, after Albert Einstein published his theory of General Relativity. At first the Universe was thought to be static: Einstein himself was convinced that the Universe *should* have been static. However, in 1929 Edwin Hubble discovered that it was *expanding*: in particular, he observed that the light coming from distant galaxies was redshifted in a way proportional to the distance of the galaxy itself. This discovery paved the way to the theory of the Big Bang, according to which the whole Universe expanded from a hot dense plasma to everything we can observe today.

The Big Bang paradigm got consolidated later in 1964, when Penzias and Wilson observed for the first time the Cosmic Microwave Background radiation (CMB). This radiation is made up of the photons that were once tightly bound with protons and electrons in the primeval plasma, when the Universe was hot and dense. With the expansion the Universe got progressively colder, until $\sim 380\,000$ years after the Big Bang it was cold enough to allow the formation of the first atoms. At this time the photons coming from the primeval plasma – not having enough energy to break the atoms – decoupled and started to free-stream across the Universe, forming what we observe today as the CMB.

Later in 1998 another important discovery in the field of Cosmology was made. From the observation of type Ia Supernovae, research groups led by Riess [1] and Perlmutter [2] obtained the first evidence that the Universe expansion is *accelerating*. This was also the first strong evidence for the possible existence of a cosmological constant (Λ), which is the simplest model of *dark energy*, a particular form of energy which can drive cosmic acceleration through its negative pressure.

These observations helped to build and establish the ΛCDM model, which nowadays is currently accepted as the Standard Model of Cosmology. According this model, about 68% of the energy content of the Universe is made up of dark energy in the form of a cosmological constant, about which we know nothing at microscopic level. The remaining 32% is made by non-relativistic matter. About 5% is ordinary (baryonic) matter made of protons, neutrons and electrons, and 27% is *dark matter*. The name "dark" comes from the fact that it does not interact via the electromagnetic force, but only through gravity. Except for this, nothing else is known about dark matter. Thus, the 95% of the Universe seems to be made of something whose physical nature is not known yet.

One of the greatest successes of the ΛCDM model is the high accuracy with which it predicts the power spectrum of the CMB temperature fluctuations. The Planck satellite [3], operating between 2009 and 2013, provided the highest resolution map of the CMB sky. The

mission data analysis [4] showed an astonishing agreement between the observations and the theoretical prediction from ΛCDM.

Despite being consistent with the observations, the ΛCDM model presents also various problems. It does not explain what dark energy and dark matter are. About the former we only know that it has a negative pressure which drives the current cosmic acceleration. About the latter we only know that it is almost non-relativistic and that it keeps galaxies and clusters of galaxies together with its gravitational pull.

In order to tackle these and other open questions, new observations are coming in the next years. One of these is *Euclid*, an ESA medium class mission which will launch a satellite by the end of 2022, and which is the context of this thesis. The *Euclid* mission will observe about one third of the sky, performing one of the largest galaxy surveys ever made and probing the last 10 billion years of the Universe expansion history.

The main cosmological probes for which *Euclid* is designed are *weak lensing* (WL) and *galaxy clustering* (GC). By weak lensing it is meant the slight deformation of the images of galaxies due to density fluctuations of the intervening matter distribution, either dark or baryonic. Through weak lensing measurements it is possible to probe the matter distribution which sources it. Galaxy clustering consists instead in studying the statistical properties of the distribution of galaxies, which is not random. In particular, it contains an oscillating pattern which is an imprint of the sound waves that were propagating in the primeval plasma of baryon and photons, which was permeating the Universe before the CMB photons decoupled. This pattern are the Baryon Acoustic Oscillations (BAO), and their length scale can be inferred from the galaxy distribution. Moreover, the galaxy distribution can also be used to *trace* the underlying dark matter field, being so-called *biased tracers* of the dark matter distribution, which cannot be observed directly.

Euclid will study these probes with two instruments: the Visible Imager (VIS) and the Near-Infrared Spectro-Photometer (NISP). With VIS *Euclid* will obtain high resolution images of 1.5 billion galaxies for weak lensing measurements. With the NISP instrument in photometric mode it will measure the photometric redshifts of the same galaxies observed with VIS. In the spectroscopic mode NISP will instead measure the spectroscopic – which are about 50 times more accurate than the photometric ones – redshifts of 20 million Hα-emitting galaxies. Thus, *Euclid* will produce two galaxy samples: a photometric sample and a spectroscopic sample. In this sense, the galaxy clustering probe can be subdivided in two: photometric galaxy clustering ($\mathrm{GC_{ph}}$) and spectroscopic galaxy clustering ($\mathrm{GC_{sp}}$). $\mathrm{GC_{sp}}$ and the so-called *Euclid* 3×2pt statistics, composed by WL, $\mathrm{GC_{ph}}$ and their cross-correlation, represent the two main probes of *Euclid*.

During the three years of my PhD at the University of Genoa, I had the opportunity to work in two complementary areas of *Euclid*. In the first half of my PhD I worked on image simulations needed for the validation of the official spectroscopic data reduction software of *Euclid*. In fact, in order to accurately measure the redshifts of the Hα-emitter galaxies of the *Euclid* spectroscopic sample, their one-dimensional spectra have to be extracted from the two-dimensional dispersed images acquired by NISP in spectroscopic mode. In this sense, a software which performs a data reduction – from images to 1D intensity-vs-wavelength spectra – is needed. In *Euclid* there is a specific Organisation Unit (OU-SIR) in charge of the development of a data reduction software, the SIR Pipeline. In order for the mission to reach the expected performances in the spectroscopic channel, the SIR Pipeline must satisfy specific requirements. Therefore, it needs to be validated before the launch with simulations of the

real data which will be acquired by NISP during the mission. The extraction of the spectra is in fact affected by several sources of systematic errors, and the effect of each of these must be accurately assessed. For this reason a wide and varied set of simulations is needed, together with full control on every detail of the simulation.

In this context I have developed a software which produces these simulations, built on top of the official *Euclid* spectroscopic simulator developed by OU-SIM. I added the possibility to easily customise all settings of a given simulation, mainly the instrumental noise effects and the input catalogue of the sources, together with their theoretical spectra. I also built a unified end-to-end pipeline which automates the run of the SIR Pipeline on the simulated images, allowing to perform a direct comparison between the input theoretical spectra and the output extracted ones. At the moment of writing, the software I have developed is being used by OU-SIR as the main tool for carrying out the validations. There are also plans for applying it to perform specific simulations for the Legacy Science groups, like the AGN and Galaxy Evolution teams.

In the second half of my PhD, as a member of the Work Package (WP) Likelihood of the Galaxy Clustering Science Working Group (GC-SWG) of *Euclid*, I performed the first *Euclid* cosmological parameter forecast which includes the correlations between $\mathrm{GC_{sp}}$ and the 3×2pt statistics. The data analysis of the survey needs in fact to be accurately planned: with this purpose pre-launch forecasts of the expected scientific performances are needed to help this planning process. In a previous *Euclid* forecast – the IST:F [5] – it has been shown that the inclusion in the analysis of the correlation between WL and $\mathrm{GC_{ph}}$ significantly improves the constraints on the cosmological parameters. In my work I extended the IST:F including also the two correlations between $\mathrm{GC_{sp}}$ and $\mathrm{GC_{ph}}$ and between $\mathrm{GC_{sp}}$ and WL, in order to understand their impact of the foreseen constraints given by *Euclid*. The results here presented show that these correlations may not affect the constraints as much as the $\mathrm{XC(WL, GC_{ph})}$ studied in the IST:F [5]. However, in the next future some extensions of my work will be studied in order to take a final decision on whether including or not these correlations in the official *Euclid* data analysis.

Thesis outline

This thesis is structured as follows.

In Chapter 1 I present the framework of Modern Cosmology. Starting from a brief recall of the homogeneous Universe, I introduce the basics for studying inhomogeneities in linear perturbation theory. Finally, I give an overview of cosmological observations and the ΛCDM model.

In Chapter 2 I give a general overview of the *Euclid* mission. I describe the scientific objectives and the consequent accuracy requirements. Then I summarise the structure of the *Euclid* spacecraft, giving also technical details about the VIS and NISP instruments. After that I describe the survey and the observational strategy, and finally I briefly present the organisation of the mission. In this last section an overview of the *Euclid* Science Ground Segment is given, along with a schematic description representation of the Processing Functions (PFs) and Organisation Units (OUs) and their tasks.

In Chapter 3 I present my work in the context of spectroscopic image simulations. I give an introduction on slitless spectroscopy and I describe the structure and requirements of the SIR Pipeline. Then I introduce the official *Euclid* simulations produced by OU-SIM, with a

focus on TIPS, the official spectroscopic simulator. Then I describe the main features of the software I have written for OU-SIR, motivating the need for the specific simulations it can produce. Finally I report some test cases in which my software was used to for validating the SIR Pipeline.

In Chapter 4 I present the forecast I have done to include the correlations between $\mathrm{GC_{sp}}$ and $\mathrm{WL}, \mathrm{GC_{ph}}$. I describe the formalism of harmonic power spectra which I have used to compute these correlations, also summarising the Fisher matrix technique in this specific context. Finally I report and discuss the results of the analysis, quantitatively assessing the impact of the aforementioned correlations on cosmological parameter constraints.

Chapter 1

Modern Cosmology

In 1929 Edwin Hubble discovered that our Universe is not static, but it is *expanding*. Hubble first observed that the light coming from distant galaxies is redshifted, which today we recognise as a sign of the wavelength stretching due to the expansion of space. In this chapter I present the fundamental concepts needed to understand the main cosmological observations, together with the ΛCDM model, which is the nowadays accepted concordance model describing the Universe.

The chapter is structured as follows: in Sec. 1.1 I describe the homogeneous and isotropic Universe, which is the starting point in Cosmology and describes the cosmos as a whole on the largest scales. In Sec. 1.2 I report the various distance definitions that are used in Cosmology, and that do not necessarily coincide between themselves since our Universe is not static. In Sec. 1.3 I present the cosmological linear perturbation theory, which is employed as a main tool for analysing cosmological data-sets. In Sec. 1.4 I briefly describe some of the most important cosmological observations that have been carried out in the years. Finally in Sec. 1.5 I summarise the characteristics and parameters of the Concordance Model of Cosmology, ΛCDM, together with its open problems and possible extensions.

Unless otherwise specified, in this chapter I am following Dodelson [6] and Carroll [7], adopting units such that

$$\hbar = 1, \quad c = 1, \quad k_B = 1\,. \tag{1.1}$$

1.1 The homogeneous and isotropic Universe

When observed at very large scales, our Universe appears homogeneous and isotropic. This is a very strong statement, and it means that the Universe looks the same to any observer, independently of the direction at which he/she is looking at and the location where he/she happens to be. This is generally stated as the *Cosmological Principle*. Moreover, at cosmological and astronomical scales the only relevant interaction is gravity, which is described by Einstein's General Theory of Relativity (GR). According to GR, space-time can be modelled as a manifold, and gravity is a manifestation of the curvature of this manifold. The geometry of space-time is encoded in the metric tensor $g_{\mu\nu}$, the solution of Einstein equations:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}\,. \tag{1.2}$$

On the left hand side there is the Einstein tensor $G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}$, with $R_{\mu\nu}$ being the Ricci tensor and R the Ricci scalar. Λ is the cosmological constant, and it was first introduced by Einstein to allow for a static Universe. On the right hand side there is the energy-momentum tensor $T_{\mu\nu}$, representing the energy content of the Universe, and G is the Newton gravitational constant. Einstein equations have a beautiful physical interpretation, given by John Wheeler:

Space-time tells matter how to move; matter tells space-time how to curve.

This sentence synthesizes the profound meaning of Einstein's theory. In other words the left hand side of Einstein equations represents the geometry of space-time, since it contains only the metric tensor and its derivatives with respect to the coordinates. The right hand side contains instead the physical energetic content of the Universe, represented by the energy-momentum tensor. An equation linking these two entities states that the space-time geometry is influenced by the mass and the energy of physical bodies embedded in it ("matter tells space-time how to curve"). In turn the bodies trajectories are influenced by space-time curved geometry ("space-time tells matter how to move").

1.1.1 The FLRW Universe

Einstein's equations are ten non-linear coupled differential equations, and for this reason they are almost always very difficult to solve. A commonly adopted strategy is to assume a particular form of the solution $g_{\mu\nu}$, depending on a few functions of the space-time coordinates, and then substitute back this form into the equations and solve for these functions. To study the Universe at large scales, the metric can be assumed to be homogeneous and isotropic. At present time, the length scale above which the Universe is homogeneous is about 100Mpc [8], which is about one hundredth of the observable Universe radius.

In the homogeneous Universe, it can be shown that the solution is the Friedman-Lemaître-Robertson-Walker (FLRW) metric [7]

$$ds^2 = -\mathrm{d}t^2 + a^2(t) \left[\frac{\mathrm{d}r^2}{1 - kr^2} + r^2 d\Omega^2 \right] . \tag{1.3}$$

Here t is the cosmic time and $a(t)$ is the scale factor, whose dynamical evolution captures the expansion of physical distances due to gravity. The scale factor conventionally is defined in such a way that $a(t_0) = 1$, with t_0 being the present cosmic time (today). The quantity k is the spatial curvature and can assume three values: -1, 0, $+1$. A FLRW Universe with $k = -1$ is named *open*; if $k = +1$ the Universe is said to be *closed*; finally a Universe with $k = 0$ is said to be *flat*. For later convenience we introduce the conformal time

$$\eta(t) \equiv \int_0^t \frac{\mathrm{d}t'}{a\,(t')} , \tag{1.4}$$

in terms of which the FLRW metric takes the form

$$\mathrm{d}s^2 = a^2(\eta) \left[-\mathrm{d}\eta^2 + \frac{\mathrm{d}r^2}{1 - kr^2} + r^2 \mathrm{d}\Omega^2 \right] . \tag{1.5}$$

Using the FLRW metric it is possible to evaluate the LHS of Einstein equations. It is also necessary to characterise the RHS through a model of the energy-momentum tensor of the

energy content of the Universe which is coherent with the hypotheses of homogeneity and isotropy leading to FLRW metric. In particular it is assumed that at very large scales, where Eq. (1.3) holds, the energy content of the Universe can be modelled [7] as a simple superposition of non-interacting *perfect fluids* in thermodynamical equilibrium:

$$T_{\mu\nu} = \sum_i T^i_{\mu\nu} = \sum_i \left[P_i g_{\mu\nu} + (P_i + \rho_i) U_\mu U_\nu \right] . \tag{1.6}$$

In the above equation P is the pressure of the fluid, ρ its energy density and U_μ its four-velocity. The perfect fluids are assumed to be at rest with respect to a reference frame which is *comoving* with the coordinates grid. In this frame the four-velocity of the fluid is simply given by

$$U^\mu = (1, 0, 0, 0) . \tag{1.7}$$

Moreover, P and ρ depend only on the cosmic time t, under the homogeneity assumption. Now substituting the FLRW metric Eq. (1.3) into the Einstein equations, together with the perfect fluid energy-momentum tensor Eq. (1.6), we get the so called Friedman equations

$$\begin{aligned} H^2 &= \frac{\Lambda}{3} + \frac{8\pi G}{3} \sum_i \rho_i - \frac{k}{a^2} \\ \dot{H} &= -4\pi G \sum_i (P_i + \rho_i) + \frac{k}{a^2} \end{aligned} \tag{1.8}$$

where we introduced the Hubble parameter $H \equiv \frac{\dot{a}}{a}$, with $H(t_0) \equiv H_0$ being called the *Hubble constant*. Another equation holding is the covariant conservation of the energy-momentum tensor

$$\nabla^\mu T_{\mu\nu} = 0 , \tag{1.9}$$

whose component with $\nu = 0$ leads to

$$\dot{\rho}_i + 3H (\rho_i + P_i) = 0 . \tag{1.10}$$

The three equations Eq. (1.10) and Eq. (1.8) are not all independent, due to the Bianchi identity

$$\nabla^\mu G_{\mu\nu} = 0 , \tag{1.11}$$

which is a purely geometrical identity holding for the metric tensor, and assures the consistency of the Einstein equations with the energy-momentum conservation Eq. (1.9).

It is physically interesting to rewrite the first Friedman equation introducing the dimensionless density parameters $\Omega_i(t)$

$$\Omega_i(t) \equiv \frac{8\pi G}{3H^2(t)} \rho_i . \tag{1.12}$$

With this definition the first Friedman equation becomes

$$\sum_i \Omega_i(t) + \frac{\Lambda}{3H^2(t)} - \frac{k}{a^2 H^2(t)} = 1 \tag{1.13}$$

It is therefore possible to treat the cosmological constant and the curvature as perfect fluids, with the following identifications

$$\begin{aligned} \Omega_\Lambda(t) = \frac{\Lambda}{3H^2(t)} &\implies \rho_\Lambda \equiv \frac{\Lambda}{8\pi G} \\ \Omega_k(t) = -\frac{k}{a^2(t)H^2(t)} &\implies \rho_k \equiv -\frac{3k}{8\pi G a^2(t)} \end{aligned} \tag{1.14}$$

And this allows us to rewrite the first Friedman equation in the compact form

$$\sum_i \Omega_i(t) = 1\,. \tag{1.15}$$

where now the sum also runs over the spatial curvature k and the cosmological constant Λ.

1.1.2 The physical components of the Universe

Within the perfect fluid model on a FLRW space-time there are three unknowns to solve for, i.e. $\rho(t), P(t), a(t)$, but only two independent equations. It is therefore necessary to impose an *equation of state*, a constitutive link between ρ and P. The equation of state clearly depends on the nature of the fluid itself: the simplest model is a linear relation between the pressure and the energy density

$$P = w\rho\,. \tag{1.16}$$

Using this equation of state, Eq. (1.10) can be solved for ρ as a function of a:

$$\rho_i = \rho_{i,0} a^{-3(1+w_i)}\,. \tag{1.17}$$

Here the subscript 0 denotes the quantity evaluated at present time t_0. Despite of its simplicity, Eq. (1.16) is able to effectively model the behavior of physically relevant perfect fluids. In what follows I list the species populating the cosmic inventory in the Standard Cosmological Model, i.e. ΛCDM.

Cold Matter Cold matter comprises all the non-relativistic components in the Universe, such as baryons and cold dark matter. These are characterised by an energy density contribution dominated by the rest energy, for which the pressure is negligible:

$$P_\mathrm{m} = 0 \implies w_\mathrm{m} = 0\,. \tag{1.18}$$

Therefore the energy density of cold matter scales as a^3 with the Universe's expansion

$$\rho_\mathrm{m} \propto a^{-3}\,. \tag{1.19}$$

This can be intuitively understood thinking the cold matter perfect fluid as a gas of massive particles in an expanding spherical box (the Universe) with radius proportional to $a(t)$. Since particles are non-relativistic, their energy is almost entirely given by their mass. Therefore the energy density is the mass volume density, which scales as the inverse of the volume of the sphere, proportional to a^3. From this reasoning Eq. (1.19) follows.

Radiation Radiation stands for relativistic particles, for which the energy is mainly given by their momentum. Radiation therefore includes photons and relativistic neutrinos. For radiation the equation of state is

$$P_r = \frac{1}{3}\rho_{\rm r}, \qquad w_r = \frac{1}{3}\,. \tag{1.20}$$

This result can be derived starting from the energy-momentum tensor of electromagnetic field

$$T^{\mu\nu}_{\rm EM} = F^{\mu\lambda}F^{\nu}{}_{\lambda} - \frac{1}{4}g^{\mu\nu}F^{\lambda\sigma}F_{\lambda\sigma}\,. \tag{1.21}$$

This tensor is trace-less $g_{\mu\nu}T^{\mu\nu}_{\rm EM} = 0$, and imposing this condition on the perfect fluid energy-momentum tensor we get the equation of state Eq. (1.20). Substituting this into Eq. (1.16) leads to

$$\rho_{\rm r} \propto a^{-4}\,. \tag{1.22}$$

For photons, this result can be explained with an argument similar to the one used for Cold Matter. In addition to the volume scaling a^{-3} of the number density photons have an energy which is inversely proportional to their wavelength, which in turn scales as $a(t)$. This explains the additional a^{-1} factor: it is simply due to a *redshift* of the photons wavelength which occurs as the Universe expands.

Cosmological Constant As we see from equation Eq. (1.14) the cosmological constant contribution is like the one of a perfect fluid with a constant energy density. Therefore from equation Eq. (1.16) it follows

$$P_\Lambda = -\rho_\Lambda\,. \tag{1.23}$$

If the energy density is imposed to be positive definite the pressure of the cosmological constant becomes *negative*. Indeed it can be seen from Eq. (1.8) that in a Universe dominated by a cosmological constant the scale factor grows exponentially

$$a(t) \propto e^{Ht}\,, \tag{1.24}$$

where H is the Hubble parameter, which is constant in a Λ-dominated Universe. The cosmological constant is the simplest model of *dark energy*, which is a still unknown physical entity composing the major part of the Universe and which pushes the accelerated expansion of the Universe with its negative pressure.

Putting together all the results obtained until here it is possible to write an evolution equation of the Hubble parameter in the homogeneous Universe with a cosmological constant

$$E^2(a) \equiv \frac{H^2(a)}{H_0^2} = \sqrt{\Omega_{\rm m,0}\,a^{-3} + \Omega_{\rm r,0}\,a^{-4} + \Omega_{\Lambda,0} + \Omega_{k,0}\,a^{-2}}\,, \tag{1.25}$$

where the notation $\Omega_{i,0} \equiv \Omega_i(t = t_0)$ has been introduced.

1.1.3 Epoch of matter-radiation equality

We have seen in the previous section that the density parameter of radiation falls of more rapidly than the density parameter of cold matter. The contribution of radiation $\Omega_{\mathrm{r},0}$ today is much smaller than the cold matter one $\Omega_{\mathrm{m},0}$, but in early Universe the situation was reversed. This means that there has been a time when the two contributions were equal, which can be quantified with the value of the scale factor a_{eq} at that time. This epoch plays an important role in the linear theory of structure formation, acting as a watershed for the behaviour of the structure evolution between small and large scales. In this section I briefly derive the value of a_{eq}, following the treatment of [6].

The majority of the radiation contribution today is given by the Cosmic Microwave Background (CMB), composed by the photons which decoupled from matter about 380 000 years after the Big Bang. The CMB can be modelled as a black-body radiation with a given temperature T, in terms of which the energy density can be written:

$$\rho_{\gamma,0} = \frac{\pi}{15} T^4 \,. \tag{1.26}$$

Combined measurements from WMAP and FIRAS give a temperature of $T = 2.72548 \pm 0.00057\mathrm{K}$ [9], from which the photon density parameter today can be estimated

$$\Omega_{\gamma,0} h^2 = 2.45 \times 10^{-5} \,. \tag{1.27}$$

Here h is the reduced Hubble parameter at present time

$$h \equiv \frac{H_0}{100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}} \,. \tag{1.28}$$

The total radiation density parameter $\Omega_{\mathrm{r},0}$ must also include the relativistic neutrinos contribution, which gives [6]

$$\frac{\rho_{\mathrm{r},0}}{\rho_{\mathrm{cr}}} = \frac{4.15 \times 10^{-5}}{h^2 a^4} \equiv \frac{\Omega_{\mathrm{r},0}}{a^4} \,. \tag{1.29}$$

The radiation density parameter today is about four order of magnitudes smaller than the cold matter one

$$\frac{\Omega_{\mathrm{r},0}}{\Omega_{\mathrm{m},0}} \simeq 3 \times 10^{-4} \tag{1.30}$$

where I have used the value $\Omega_{\mathrm{m},0} = 0.32$ as measured by Planck [4]. Equating the RHS of the above equation to the matter density parameter $\Omega_{\mathrm{m},0}/a^3$ yields the value of the scale factor at the epoch of matter-radiation equality

$$a_{\mathrm{eq}} = \frac{4.15 \times 10^{-5}}{\Omega_{\mathrm{m},0}\, h^2} \,. \tag{1.31}$$

This can be translated into a value for the redshift z_{eq} at the epoch of equality

$$1 + z_{\mathrm{eq}} = 2.38 \times 10^4\, \Omega_{\mathrm{m},0}\, h^2 \,. \tag{1.32}$$

To conclude this section we note that if the amount of matter $\Omega_{\mathrm{m},0} h^2$ in the Universe today goes up, the redshift at equality z_{eq} also goes up. That is, the epoch of matter-radiation equality happens earlier in a Universe with a larger matter content.

1.2 Distances

Measuring distances in Cosmology requires particular attention to the definition of the distance itself. In fact the *physical distance* between two objects increases with time because of the Universe expansion, the so-called Hubble flow. All the distances definition in Cosmology can be related to the *comoving* distance, which is defined to be unaffected by the Hubble flow, and therefore remains fixed as the Universe expands. Also, since the physical distances depend on the cosmological model, it is necessary to specify the cosmology within which to compute and interpret them.

1.2.1 Redshift

The light coming from a distant source – e.g. a galaxy – undergoes a wavelength shift due to the stretching of the three-dimensional space encoded in the scale factor $a(t)$. Defining $t_{\rm em}$ and $t_{\rm obs}$ the cosmic times of emission and detection of the light signal respectively, the ratio of the wavelength at the two different times is given by

$$\frac{\lambda_{\rm obs}}{\lambda_{\rm em}} = \frac{a(t_{\rm obs})}{a(t_{\rm em})} > 1\,. \tag{1.33}$$

Since the time of observation is the present time t_0 and conventionally the scale factor is normalised such as $a(t_0) = 1$, the *redshift* z of the emitted light is given by

$$z \equiv \frac{\lambda_{\rm obs} - \lambda_{\rm em}}{\lambda_{\rm em}} = \frac{1}{a} - 1\,, \tag{1.34}$$

where $a \equiv a(t_{\rm em})$ is the scale factor at the generic time of emission. The above equation can be inverted to relate a to z

$$a = \frac{1}{1+z}\,. \tag{1.35}$$

The redshift z of a celestial object can be directly measured, and its interpretation does not depend on the assumed cosmological model. Instead the scale factor $a(t)$ appearing in Eq. (1.3) is the solution of the Friedman equations Eq. (1.8), therefore its functional dependence on the cosmic time t is affected by the cosmological model. For this reason z is used as the preferred "time variable" in Cosmology, using the one-to-one correspondence Eq. (1.35) to do the appropriate transformations in the equations.

The *Euclid* mission, which is the context of this thesis work, will perform a galaxy redshift survey of unprecedented size, covering an area of about one third of the sky in the redshift range $0.001 < z < 2.5$.

1.2.2 Comoving distance

The comoving distance quantifies the distance between two objects with respect to the comoving coordinate grid (r, θ, ϕ). This coordinate grid expands with the Universe, and therefore is unaltered by the Universe expansion. If two objects are at rest with respect to each other they are said to have zero peculiar relative velocities, and their comoving distance stays fixed. On the other hand their physical distance increases solely because of the Hubble flow. The

comoving distance between a distant source emitting light at time t and us is defined as

$$\chi \equiv \int_t^{t_0} \frac{dt'}{a\left(t'\right)} = \int_{a(t)}^1 \frac{da'}{a'^2 H\left(a'\right)} = \int_0^z \frac{dz'}{H\left(z'\right)} \,. \tag{1.36}$$

Physically the comoving distance between two objects is the maximum distance which can be travelled by light when going from one object to the other. In particular the comoving distance that light could have travelled from $t = 0$ is referred as the *comoving horizon*, and it is given by the conformal time defined in Eq. (1.4). Since no information can exceed the speed of light, two events separated by a comoving distance greater than η cannot be causally related to each other.

1.2.3 Angular diameter distance

Despite being useful from a conceptual point of view, the comoving distance cannot be observed directly. Thus cosmologists had to find other ways to define observable distances. A possible way to determine distances is to measure the angular size θ of an object of known physical size l. Since distances in Cosmology are always much bigger than the physical sizes of the observed objects, the angular size θ of these objects is small. With this in mind the angular diameter distance can be defined as

$$d_A = \frac{l}{\theta} \,. \tag{1.37}$$

The comoving size of the object in a *flat* ($\Omega_{k,0} = 0$) is l/a so the angle subtended can be computed as $\theta = \frac{l/a}{\chi}$, thus producing

$$d_A = a\chi = \frac{\chi(z)}{1+z} \,. \tag{1.38}$$

For a non-flat Universe the above expression can be generalised as

$$d_A(z) = \frac{1}{(1+z)H_0\sqrt{|\Omega_{k,0}|}} S_k\left(H_0\sqrt{|\Omega_{k,0}|}\chi(z)\right) \tag{1.39}$$

where S_k is a function whose expression varies depending on the spatial curvature k of the FLRW metric Eq. (1.3):

$$S_k(\chi) = \begin{cases} \sin\chi & k = +1 \\ \chi & k = 0 \\ \sinh\chi & k = -1 \end{cases} \tag{1.40}$$

The angular diameter distance is used for example in measurements of gravitational lensed quasar systems [10].

1.2.4 Luminosity distance

Another way to infer distances is to measure the light flux F coming from a distant source of known luminosity, i.e. a standard candle. In a static Universe the flux F emitted from a source of luminosity L located at distance d from the observer is

$$F = \frac{L}{4\pi d^2} \,. \tag{1.41}$$

This equation can be generalised [6] to the expanding Universe using the comoving distance

$$F = \frac{La^2}{4\pi\chi^2(a)}\,, \tag{1.42}$$

leading to the expression for the luminosity distance d_L as a function of the redshift

$$d_L(z) = (1+z)\chi(z)\,. \tag{1.43}$$

The above definition holds only for the flat case. In a non-flat FLRW cosmology the formula generalises to

$$d_L(z) = \frac{(1+z)}{H_0\sqrt{|\Omega_{k,0}|}} S_k\left[\frac{\sqrt{|\Omega_{k,0}|}}{H_0}\chi(z)\right]\,, \tag{1.44}$$

The luminosity distance $d_L(z)$ can be measured, as well as the redshift z. This allows to fit the experimental data with the relation Eq. (1.44) and therefore infer the best fit values of the cosmological parameters. This has been done with Supernovae Ia, as reported in Sec. 1.4.2.

1.3 The inhomogeneous Universe

Our Universe appears homogeneous at very large scales, but as soon as we start zooming in at smaller scales inhomogeneities arise. A representation of the results of the Sloan Digital Sky Survey (SDSS) [11] is displayed in Fig. 1.1; each dot is a galaxy, and the maximum redshift represented is $z \sim 0.15$. The figure clearly shows that the distribution of the galaxies is not uniform, but there are overdense and underdense spots. The inhomogeneities in the matter distribution are the result of the evolution history of the Universe, starting from inflation up to the present day. This evolution depends on the values of cosmological parameters, and therefore there is information in the structure of the inhomogeneous Universe. Galaxy surveys, like SDSS and Euclid, are a way to probe this structure, as they are aimed at mapping the distribution of galaxies. In turn the galaxies *trace* the distribution of dark matter, in the sense that they aggregate along the filaments of the cosmic web made by dark matter density fluctuations. Therefore galaxy surveys can be used to infer the properties of the cold matter distribution; more on galaxy survey will be discussed in Sec. 1.4.3. The equations controlling the evolution of the structures in the Universe are coupled Einstein-Boltzmann equations. In the regime where the inhomogeneities are assumed to be small one can keep only the linear leading terms in the equations. In this section I briefly review the linear theory leading to the relevant equations for growth of structures. The starting point is the perturbed FLRW metric

$$\begin{aligned} g_{00}(\boldsymbol{x},t) &= -1 - 2\Psi(\boldsymbol{x},t) \\ g_{0i}(\boldsymbol{x},t) &= 0 \\ g_{ij}(\boldsymbol{x},t) &= a^2(t)\delta_{ij}[1+2\Phi(\boldsymbol{x},t)]\,. \end{aligned} \tag{1.45}$$

The functions Ψ and Φ are the gravitational potential and the curvature perturbation respectively, and both are assumed to be small in linear perturbation theory. This expression for the metric corresponds to a particular gauge choice, i.e. the conformal Newton gauge, and contains only scalar perturbations. However, since General Relativity possesses a gauge freedom, different gauge choices describe the same physics, and thus the calculations can be carried out after having fixed the gauge.

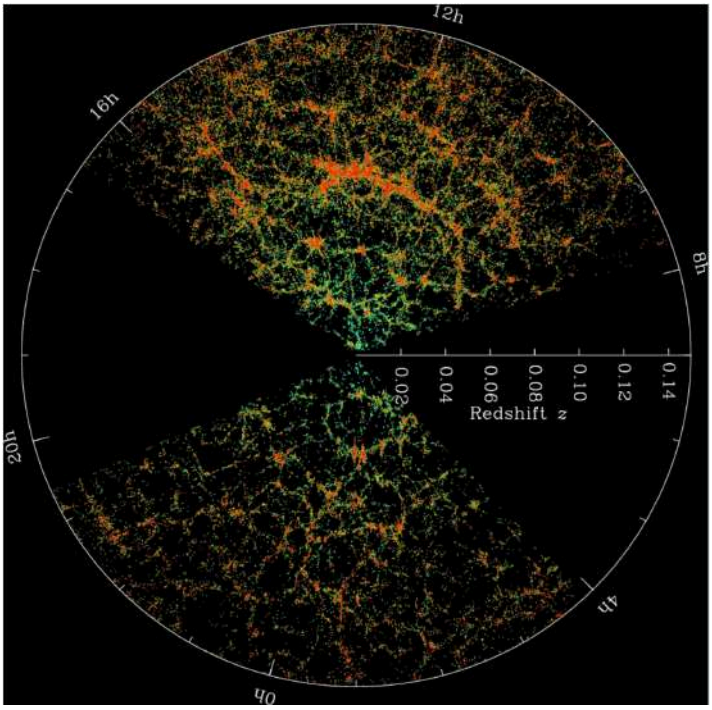

Figure 1.1: The SDSS's map of the Universe. Each dot is a galaxy; the color is the g-r color of that galaxy. Link: `https://www.sdss.org/science/orangepie/`

1.3.1 Boltzmann equations

Theoretical Cosmology is essentially an application of general relativity coupled with statistical mechanics [6]. The only relevant long-range force is gravity, governed by Einstein equations Eq. (1.2). Since Cosmology deals with the evolution of the Universe on large scales, it cares about the collective, average behavior of matter, which is described by statistical mechanics. The main equation of statistical mechanics is the *Boltzmann equation*, and its unknown is the phase-space distribution function $f(\boldsymbol{x}, \boldsymbol{p}, t)$. Given a small volume $(\Delta x \Delta p)^3$ centred at $(\boldsymbol{x}, \boldsymbol{p})$ in phase-space, the distribution function gives the number of particles enclosed in this volume

$$N(\boldsymbol{x}, \boldsymbol{p}, t) = f(\boldsymbol{x}, \boldsymbol{p}, t) \left(\frac{\Delta x \Delta p}{2\pi} \right)^3 . \tag{1.46}$$

The generic form of the Boltzmann equation is the following

$$\frac{df}{dt} = C[f] \, . \tag{1.47}$$

Essentially it encodes the conservation of the number of particles, considering also annihilation and creation of particles in the relativistic limit with the collision term $C[f]$. The derivative on the LHS is a total derivative, which can expanded as

$$\frac{\partial f}{\partial t} + \frac{\mathrm{d}x^i}{\mathrm{d}t} \frac{\partial f}{\partial x^i} + \frac{dp}{\mathrm{d}t} \frac{\partial f}{\partial p} + \frac{d\hat{p}^i}{\mathrm{d}t} \frac{\partial f}{\partial \hat{p}^i} = C[f] \, . \tag{1.48}$$

Here p indicates the module of the momentum of the particles, while $\hat{p}^i$ is a unit vector which represents the direction of propagation of the particles. The Boltzmann equation has to be complemented with the equations of motion for the particles. In GR these are summarised in the geodesic equation

$$\frac{\mathrm{d}P^\mu}{\mathrm{d}\lambda} + \Gamma^\mu_{\alpha\beta} P^\alpha P^\beta = 0\,, \tag{1.49}$$

where $P_\mu \equiv \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}$ is the comoving momentum, i.e. the momentum in the comoving reference frame. Once the Boltzmann equation for a given species are solved, macroscopic quantities can be computed. Macroscopic quantities include the energy momentum tensor

$$T^\mu{}_\nu\,(\boldsymbol{x}, t) = \frac{g_s}{\sqrt{-\det\left[g_{\alpha\beta}\right]}} \int \frac{dP_1 dP_2 dP_3}{(2\pi)^3} \frac{P^\mu P_\nu}{P^0} f(\boldsymbol{x}, \boldsymbol{p}, t)\,, \tag{1.50}$$

and the pressure

$$P(\boldsymbol{x}, t) = g_s \int \frac{d^3p}{(2\pi)^3} f(\boldsymbol{x}, \boldsymbol{p}, t) \frac{p^2}{3E(p)}\,. \tag{1.51}$$

For both quantities the factor g_s represents the quantum degeneracy factor of the species considered, which is typically given by the number of spin (or helicity) states. In what follows I briefly report the Boltzmann equation for all the species of particle relevant for the expansion history of the Universe.

Photons

As anticipated in Sec. 1.1.3, the majority of the radiation in our Universe is composed by the CMB photons. At zero order the photons perfect fluid can be modelled as a Bose-Einstein gas with corresponding equilibrium distribution function

$$f^{(0)}(p) = \frac{1}{e^{p/T} - 1}\,. \tag{1.52}$$

From this Eq. (1.26) can be derived by integrating the photons energy against $f^{(0)}$

$$\rho_\gamma = 2 \int \frac{d^3p}{(2\pi)^3} \frac{p}{e^{p/T} - 1} = \frac{8\pi T^4}{(2\pi)^3} \int_0^\infty dx \frac{x^3}{e^x - 1} = \frac{\pi}{15} T^4\,, \tag{1.53}$$

where the dimensionless variable $x \equiv p/T$ is introduced. Beyond the zeroth order the equilibrium distribution function is slightly perturbed, and this can be done by introducing the relative temperature fluctuation $\Theta(\boldsymbol{x}, \hat{\boldsymbol{p}}, t)$

$$\Theta(\boldsymbol{x}, \hat{\boldsymbol{p}}, t) \equiv \frac{\delta T(\boldsymbol{x}, \hat{\boldsymbol{p}}, t)}{T(t)}\,. \tag{1.54}$$

With this definition the perturbed distribution function takes the following form

$$f(\boldsymbol{x}, p, \hat{\boldsymbol{p}}, t) = \left[\exp\left\{\frac{p}{T(t)[1 + \Theta(\boldsymbol{x}, \hat{\boldsymbol{p}}, t)]}\right\} - 1\right]^{-1}\,. \tag{1.55}$$

Note that here the temperature perturbation depends also on the position $\boldsymbol{x}$ in space and the direction of propagation $\hat{\boldsymbol{p}}$. This goes in the direction of modelling the CMB perturbation on

the sky observed by Earth today, which can be expressed as $\Theta\left(\boldsymbol{x}_{\rm Earth}\,,\hat{\boldsymbol{p}},t_0\right)$. At first order in Θ the distribution function can be expanded as

$$f(\boldsymbol{x},\boldsymbol{p},t)\simeq f^{(0)}(p,t)-p\frac{\partial f^{(0)}(p,t)}{\partial p}\Theta(\boldsymbol{x},\hat{\boldsymbol{p}},t)\,. \tag{1.56}$$

Plugging in this expansion together with the geodesic equation Eq. (1.49) at linear order into Eq. (1.48) leads to the left hand side of the Boltzmann equation for photons

$$\frac{df}{dt}=-p\frac{\partial f^{(0)}}{\partial p}\left[\dot{\Theta}+\frac{\hat{p}^i}{a}\frac{\partial\Theta}{\partial x^i}+\dot{\Phi}+\frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i}\right]. \tag{1.57}$$

The other physical ingredient needed for writing down the Boltzmann equation for photons is a model for the collision term. Before the decoupling, the photons were interacting with the electrons through Compton scattering

$$e^-(\boldsymbol{q})+\gamma(\boldsymbol{p})\leftrightarrow e^-\left(\boldsymbol{q}'\right)+\gamma\left(\boldsymbol{p}'\right)\,. \tag{1.58}$$

The motion of the electrons can be treated as non-relativistic and with the same bulk velocity $\boldsymbol{u}_b$ of the baryons. It can be shown [6] that the collision term due to Compton scattering is given by

$$C[f(\boldsymbol{p})]=-p\frac{\partial f^{(0)}}{\partial p}n_e\sigma_{\rm T}\left[\Theta_0-\Theta(\hat{\boldsymbol{p}})+\hat{\boldsymbol{p}}\cdot\boldsymbol{u}_{\rm b}\right], \tag{1.59}$$

where n_e is the electron number density, $\sigma_{\rm T}=0.665\times10^{-24}\,{\rm cm}^2$ is Thomson cross-section and Θ_0 is the monopole moment of the temperature perturbation

$$\Theta_0(\boldsymbol{x},t)\equiv\frac{1}{4\pi}\int d\Omega'\Theta\left(\hat{\boldsymbol{p}}',\boldsymbol{x},t\right). \tag{1.60}$$

It is convenient to adopt the conformal time η as temporal variable and to represent equations in Fourier space. The fluid velocity field of the baryons $\boldsymbol{u}_{\rm b}$ is assumed to be irrotational, which in Fourier space means

$$u_b^i=\left(k^i/k\right)u_{\rm b}\,. \tag{1.61}$$

With these assumptions, the photon perturbation equation becomes

$$\Theta'+ik\mu\Theta+\Phi'+ik\mu\Psi=-\tau'\left[\Theta_0-\Theta+\mu u_{\rm b}\right]. \tag{1.62}$$

Here the prime denotes differentiation w.r.t. η, $\mu\equiv\frac{\boldsymbol{k}\cdot\hat{\boldsymbol{p}}}{k}$ is the cosine of the angle between the wavevector $\boldsymbol{k}$ and the photon direction $\hat{\boldsymbol{p}}$, while τ is the optical depth

$$\tau(\eta)\equiv\int_\eta^{\eta_0}d\eta' n_e\sigma_{\rm T}a,\qquad \tau'\equiv\frac{d\tau}{d\eta}=-n_e\sigma_{\rm T}a\,. \tag{1.63}$$

The optical depth quantifies the opacity of the electron gas to photons, and so it relates to the photons mean free path. At late times, the free electron density is small, so $\tau\ll1$, while at early times it is very large, then $\tau\gg1$. In Eq. (1.59) the angular dependency of the cross-section has been neglected; if considered, it finally leads to the Boltzmann equation for the photon perturbation

$$\Theta'+ik\mu\Theta=-\Phi'-ik\mu\Psi-\tau'\left[\Theta_0-\Theta+\mu u_{\rm b}-\frac{1}{2}\mathcal{P}_2(\mu)\Theta_2\right]. \tag{1.64}$$

In the above equation $\mathcal{P}_l(\mu)$ is the l-th Legendre polynomial and Θ_l is the l-th multipole moment of Θ

$$\Theta_l(k,\eta) \equiv \frac{1}{(-i)^l}\int_{-1}^{1}\frac{d\mu}{2}\mathcal{P}_l(\mu)\Theta(\mu,k,\eta)\,. \tag{1.65}$$

Cold dark matter

In all viable models of structure formation, cold dark matter (CDM) plays an important role in the growth of structure through its gravitational pull. Dark matter does not directly interact with any of the other constituents in the universe, so there is no collision term in the Boltzmann equation. Moreover its motion is non-relativistic, therefore the velocity can be treated as a perturbation. Proceeding as for the photons, the LHS of Boltzmann equation for non relativistic particles is

$$\frac{\partial f_{\rm c}}{\partial t} + \frac{p}{E}\frac{\hat{p}^i}{a}\frac{\partial f_{\rm c}}{\partial x^i} - \left[H + \dot{\Phi} + \frac{E}{ap}\hat{p}^i\Psi_{,i}\right]p\frac{\partial f_{\rm c}}{\partial p} = 0\,. \tag{1.66}$$

The calculation then proceeds taking the first moments of the distribution function, which are the number density

$$n_{\rm c} = \int\frac{d^3p}{(2\pi)^3}f_{\rm c}\,, \tag{1.67}$$

and the fluid bulk velocity

$$u^i_{\rm c} \equiv \frac{1}{n_{\rm c}}\int\frac{d^3p}{(2\pi)^3}f_{\rm c}\frac{p\hat{p}^i}{E(p)}\,. \tag{1.68}$$

Taking the first moment of Eq. (1.66) yields

$$\frac{\partial n_{\rm c}}{\partial t} + \frac{1}{a}\frac{\partial\left(n_{\rm c}u^i_{\rm c}\right)}{\partial x^i} + 3[H+\dot{\Phi}]n_{\rm c} = 0\,. \tag{1.69}$$

The CDM density contrast is introduced as a fluctuation from the homogeneous background number density

$$n_{\rm c}(\boldsymbol{x},t) = \bar{n}_{\rm c}(t)\left[1+\delta_{\rm c}(\boldsymbol{x},t)\right]\,. \tag{1.70}$$

Inserting this definition in Eq. (1.69) produces a zero order term and a first order term, which should vanish separately. The equation for the zero order term is the known evolution Eq. (1.19) of the background cold matter density

$$\frac{\partial\bar{n}_{\rm c}}{\partial t} + 3H\bar{n}_{\rm c} = \frac{d\left(\bar{n}_{\rm c}a^3\right)}{dt} = 0 \Rightarrow \bar{n}_{\rm c}\propto a^{-3}\,. \tag{1.71}$$

The first order term instead leads to an equation for the density contrast $\delta_{\rm c}$

$$\frac{\partial\delta_{\rm c}}{\partial t} + \frac{1}{a}\frac{\partial u^i_{\rm c}}{\partial x^i} + 3\dot{\Phi} = 0\,. \tag{1.72}$$

Taking the second moment of Eq. (1.66) an equation for the fluid bulk velocity $\boldsymbol{u}_{\rm c}$ can be obtained

$$\frac{\partial u^j_{\rm c}}{\partial t} + Hu^j_{\rm c} + \frac{1}{a}\frac{\partial\Psi}{\partial x^j} = 0\,. \tag{1.73}$$

The fluid bulk velocity $\boldsymbol{u}_{\rm c}$ is assumed to be irrotational as the baryon velocity field

$$u^i_{\rm c} = \left(k^i/k\right) u_{\rm c}\,. \tag{1.74}$$

Going to Fourier space and using η as time variable the Boltzmann equations for CDM finally become

$$\delta'_{\rm c} + iku_{\rm c} + 3\Phi' = 0 \tag{1.75}$$

$$u'_{\rm c} + \frac{a'}{a}u_{\rm c} + ik\Psi = 0\,. \tag{1.76}$$

Baryons

The next component of the universe that requires a set of Boltzmann equations are the electrons and protons. These components are often grouped together and called baryons. Electrons and protons are coupled by Coulomb scattering, and the rate of this interaction is much larger than the Hubble parameter at each epoch of interest. This forces the electron and proton density contrast to a common value

$$\frac{\rho_e - \bar{\rho}_e}{\bar{\rho}_e} = \frac{\rho_p - \bar{\rho}_p}{\bar{\rho}_p} \equiv \delta_{\rm b}\,, \tag{1.77}$$

and analogously for the fluid velocities

$$\boldsymbol{u}_e = \boldsymbol{u}_p \equiv \boldsymbol{u}_{\rm b}\,. \tag{1.78}$$

The procedure of taking moments of the LHS of Boltzmann equation to derive equations for δ_b and $\boldsymbol{u}_b$ proceeds just as in the case of CDM. However for baryons the Compton scattering with photons must be taken into account. The equation coming from the first moment of Boltzmann equation is completely analogous to the first equation of Eq. (1.75) for CDM

$$\delta'_{\rm b} + iku_{\rm b} + 3\Phi' = 0\,. \tag{1.79}$$

This is because Compton scattering Eq. (1.58) preserves the number of particle, and this cannot affect the first moment of Boltzmann equation, which is the continuity equation from a fluid dynamics perspective. The first moment contains a collision term on the RHS instead

$$\frac{\partial u^j_{\rm b}}{\partial t} + Hu^j_{\rm b} + \frac{1}{a}\frac{\partial \Psi}{\partial x^j} = \frac{1}{\rho_{\rm b}}F^j_{e\gamma}(\boldsymbol{x}, t)\,. \tag{1.80}$$

The second Boltzmann equation for baryons is obtained by evaluating the collision term, going to Fourier space and employing the conformal time as usual

$$u'_{\rm b} + \frac{a'}{a}u_{\rm b} + ik\Psi = \tau'\frac{4\rho_\gamma}{3\rho_{\rm b}}\left[3i\Theta_1 + u_{\rm b}\right]\,. \tag{1.81}$$

Neutrinos

Finally, we turn to the remaining abundant species of particles, the neutrinos, with distribution function $f_\nu(\boldsymbol{x}, \boldsymbol{p}, t)$. The procedure is analogous to the case of photons, since neutrinos are relativistic in the early Universe. Neutrinos are fermions, so at equilibrium they follow the Fermi-Dirac distribution function with temperature $T_\nu(a)$

$$f_\nu^{(0)}(p) = \frac{1}{e^{\frac{p}{T_\nu}} + 1} \,. \tag{1.82}$$

It can be shown [6] that at zero order the temperature of the neutrino gas is related to the the one of the photon gas by a simple constant

$$\frac{T_\nu}{T} = \left(\frac{4}{11}\right)^{1/3} \,. \tag{1.83}$$

The perturbed neutrino distribution function is instead

$$\begin{aligned} f_\nu(\boldsymbol{x}, \boldsymbol{p}, t) &= \left[\exp\left\{\frac{p}{T_v(t)[1 + \mathcal{N}(\boldsymbol{x}, \hat{\boldsymbol{p}}, t)]}\right\} + 1\right]^{-1} \\ &= \left[1 - \mathcal{N}(\boldsymbol{x}, \boldsymbol{p}, t) p \frac{d}{dp}\right] f_\nu^{(0)}(p) \,, \end{aligned} \tag{1.84}$$

where $\mathcal{N}(\boldsymbol{x}, \boldsymbol{p}, t)$ the neutrino temperature fluctuation. It is well known from the observed neutrino oscillations [12] that neutrinos are massive particles. Moreover during the epochs of interest, that is, from neutrino decoupling onward, any non-gravitational interactions of neutrinos are negligible. Thus the appropriate Boltzmann equation is the collision-less Boltzmann equation for massive particles

$$\frac{df_\nu}{dt} = \frac{\partial f_\nu}{\partial t} + \frac{p}{E_\nu(p)} \frac{\hat{p}^i}{a} \frac{\partial f_\nu}{\partial x^i} - \left[H + \dot{\Phi} + \frac{E_\nu(p)}{ap} \hat{p}^i \Psi_{,i}\right] p \frac{\partial f_\nu}{\partial p} = 0 \,. \tag{1.85}$$

Inserting the form of the perturbed neutrino distribution function and switching to Fourier space and conformal time

$$\mathcal{N}'(\boldsymbol{k}, p, \mu, \eta) + ik\mu \frac{p}{E_\nu(p)} \mathcal{N} - Hp \frac{\partial}{\partial p} \mathcal{N} = -\Phi' - ik\mu \frac{E_\nu(p)}{p} \Psi \,, \tag{1.86}$$

which is the first-order Boltzmann equation for neutrinos. It resembles the equation for photons Eq. (1.64) apart from the absent collision term, and some factors of $p/E_\nu(p)$ which becomes 1 when the neutrinos are relativistic.

Massive neutrinos Neutrinos are massive particles and there are three neutrino mass eigenstates. A recent global analysis [13] of neutrino oscillations experiments gave the following results for the mass-splittings

$$\begin{aligned} \Delta m^2_{21} &= 7.39^{+0.21}_{-0.20} \times 10^{-5}\,\mathrm{eV}^2 \\ \Delta m^2_{31} &= 2.525^{+0.033}_{-0.032} \times 10^{-3}\,\mathrm{eV}^2 \text{ (NH)} \\ \Delta m^2_{32} &= -2.512^{+0.034}_{-0.032} \times 10^{-3}\,\mathrm{eV}^2 \text{ (IH)} \end{aligned} \,, \tag{1.87}$$

with NH and IH stand for normal hierarchy ($m_1 < m_2 < m_3$) and inverted hierarchy ($m_3 < m_1 < m_2$) respectively. At least two neutrino states have a large enough mass for being non-relativistic today, thus making up a small fraction of the dark matter of the Universe. For this reason the more massive neutrinos contribute to the cold matter budget through a density parameter

$$\Omega_{\nu,0} \equiv \frac{\sum m_\nu}{93.14\, h^2}\,. \tag{1.88}$$

Thus the first way in which massive neutrinos affect the growth of structures is entering the Friedman equation Eq. (1.25) for the Hubble parameter, increasing the contribution of cold matter $\Omega_{\rm m,0}$. The second effect of the mass of neutrinos is to damp the growth of small-scale structures. In fact neutrinos move fast, as they are not cold dark matter, and free-stream out of high-density regions. Perturbations on scales smaller than the free-streaming scale are therefore suppressed. An estimate of the neutrino free-streaming scale is given [6] by

$$k_{\rm fs}(a) \simeq 0.063\, h{\rm Mpc}^{-1} \frac{\sum m_\nu}{0.1{\rm eV}} \frac{a^2 H(a)}{H_0}\,. \tag{1.89}$$

The Katrin tritium decay experiment [14] provided the current estimate on the upper bound for the absolute neutrino mass scale, yielding a limit of 1.1 eV at 90% CL.

1.3.2 Einstein equations

Boltzmann equations encode how the distribution of matter-energy is influenced by interactions, after having modelled the collision term, and gravity, through the geodesic equation. The influence of matter-energy on the metric is instead contained in the Einstein equations Eq. (1.2), which can be expanded at linear order in the metric perturbations Ψ and Φ. The procedure is straightforward, but computationally laborious. The Einstein tensor is defined as

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}\,, \tag{1.90}$$

where $R_{\mu\nu}$ is the Ricci tensor and $R = g^{\mu\nu} R_{\mu\nu}$ is the Ricci scalar. The Ricci tensor is expressed as

$$R_{\mu\nu} = \partial_\rho \Gamma^\rho_{\mu\nu} - \partial_\nu \Gamma^\rho_{\rho\mu} + \Gamma^\rho_{\rho\lambda}\Gamma^\lambda_{\mu\nu} - \Gamma^\lambda_{\rho\mu}\Gamma^\rho_{\lambda\nu}\,, \tag{1.91}$$

where $\Gamma^\mu_{\rho\sigma}$ is the Christoffel symbol, which is in turn given by

$$\Gamma^\mu_{\rho\sigma} = \frac{1}{2} g^{\mu\alpha} \left(\partial_\sigma g_{\rho\alpha} + \partial_\rho g_{\sigma\alpha} - \partial_\lambda g_{\rho\sigma}\right)\,. \tag{1.92}$$

Inserting the metric Eq. (1.45) into Einstein equations gives rise to a zero order term and a first order one. The zero order term is simply given by Friedman equations Eq. (1.8) with $k = 0$, while the first order term is the one we need here. Two equations are needed, since the metric perturbations are two. The first equation can be obtained by considering the 00 component

$$\delta G^0{}_0 = 8\pi G\, \delta T^0{}_0 \tag{1.93}$$

The Einstein tensor in Fourier space at linear order is

$$\delta G^0{}_0 = -6H\Phi_{,0} + 6\Psi H^2 - 2\frac{k^2\Phi}{a^2}\,, \tag{1.94}$$

where the comma notation stands for the derivative $\Phi_{,\alpha} \equiv \frac{\partial \Phi}{\partial x^\alpha}$. For the energy-momentum tensor the 00 component is the total energy density, and inserting it and switching to Fourier space and conformal time produces the first equation

$$k^2\Phi + 3\frac{a'}{a}\left(\Phi' - \Psi\frac{a'}{a}\right) = 4\pi G a^2 \left[\rho_{\rm c}\delta_{\rm c} + \rho_{\rm b}\delta_{\rm b} + 4\rho_\gamma\Theta_0 + 4\rho_\nu\mathcal{N}_0\right] . \tag{1.95}$$

The other equation can be obtained by projecting the purely spatial ij component with the tensor $\left(\hat{k}_i\hat{k}^j - \frac{1}{3}\delta_i^j\right)$ which gives

$$k^2(\Phi + \Psi) = -32\pi G a^2 \left[\rho_\gamma\Theta_2 + \rho_\nu\mathcal{N}_2\right] . \tag{1.96}$$

This last equation tells that the two gravitational potentials Φ and Ψ are equal and opposite, unless the photons or neutrinos exhibit a non-negligible quadrupole moment.

1.3.3 Initial conditions

The current most satisfying hypothesis about what generated the initial conditions for the cosmological evolutionary equations is *inflation*. The term "inflation" indicates a period of exponentially rapid expansion of the Universe, at the end of which the Universe expanded by many orders of magnitude. Originally inflation was introduced to explain the horizon and the flatness problem [15, 16]. In a few words, these problems consist in asking the reason why the Universe is almost spatially flat and so homogeneous, even at scales that cannot have been in causal contact. One of the experimental measurements of the Universe smoothness is the CMB temperature fluctuation, which is treated briefly later in this chapter. Introductory reviews about inflation can be found at [17] or at [6] chapter 7. The initial conditions set up by inflation can be obtained by studying the Einstein-Boltzmann equations at early times, when $k\eta \ll 1$. In this limit, the Boltzmann equations for radiation and cold matter become

$$\begin{aligned} &\Theta_0' + \Phi' = 0 \\ &\mathcal{N}_0' + \Phi' = 0 \\ &\delta_{\rm c}' = -3\Phi' \\ &\delta_{\rm b}' = -3\Phi' . \end{aligned} \tag{1.97}$$

For Einstein equations, at early times radiation dominates; the 0-0 component Eq. (1.95) becomes

$$3\frac{a'}{a}\left(\Phi' - \frac{a'}{a}\Psi\right) = 16\pi G a^2 \left(\rho_\gamma\Theta_0 + \rho_\nu\mathcal{N}_0\right) . \tag{1.98}$$

Combining it with the early times Boltzmann equation for radiation and the second Einstein equation Eq. (1.96) and neglecting the quadrupole moment[1] yields

$$\Phi''\eta + 4\Phi' = 0 . \tag{1.99}$$

This equation can be solved with the ansatz [6] $\Phi = \eta^p$, which produces two solutions: $p = 0$ or $p = -3$. The $p = -3$ mode is the decaying mode: if it is excited very early on, it will

[1] Neglecting the quadrupole moment means that $\Psi = -\Phi$ everywhere in the equations.

quickly die out and have no impact on the Universe. The interesting solution is therefore the one with $p = 0$, which inserted in Eq. (1.98) produces

$$\Phi = 2\Theta_{\mathrm{r},0}\,. \tag{1.100}$$

To go ahead the condition of *adiabatic perturbations* must be assumed:

$$\frac{\delta\rho_s}{\rho_s} = \frac{\delta\rho}{\rho} \tag{1.101}$$

which means that at early times the fractional density perturbations are the same for all species. This hypothesis has been tested with the Planck mission [4]. Inserting this condition in the Einstein Boltzmann equations leads to the following initial conditions for the overdensities and the radiation monopoles

$$\Phi\left(\boldsymbol{k},\eta_i\right) = 2\Theta_0\left(\boldsymbol{k},\eta_i\right) = 2\mathcal{N}_0\left(\boldsymbol{k},\eta_i\right)\,, \tag{1.102}$$

and for the cold matter overdensities

$$\delta_{\mathrm{c}}\left(\boldsymbol{k},\eta_i\right) = \delta_{\mathrm{b}}\left(\boldsymbol{k},\eta_i\right) = 3\Theta_0\left(\boldsymbol{k},\eta_i\right) = \frac{3}{2}\Phi\left(\boldsymbol{k},\eta_i\right)\,, \tag{1.103}$$

where η_i is the initial conformal time. For the fluid velocities of cold matter and dipole moments of radiation analogous relations hold

$$\Theta_1(\boldsymbol{k},\eta_i) = \mathcal{N}_1(\boldsymbol{k},\eta_i) = \frac{iu_{\mathrm{b}}(\boldsymbol{k},\eta_i)}{3} = \frac{iu_{\mathrm{c}}(\boldsymbol{k},\eta_i)}{3} = -\frac{k}{6aH}\Phi(\boldsymbol{k},\eta_i)\,. \tag{1.104}$$

It is remarkable that the adiabatic condition Eq. (1.101) allows to obtain the initial conditions on all perturbations from the value of the gravitational potential Φ alone. It is commonly assumed that the primordial fluctuations of the field Φ_p are gaussian with zero mean $\langle\Phi_p(\boldsymbol{k})\rangle = 0$ and with given two-point function

$$\left\langle\Phi_p(\boldsymbol{k})\Phi_p^*\left(\boldsymbol{k}'\right)\right\rangle = P_{\Phi_p}(k)(2\pi)^3\delta^{(3)}\left(\boldsymbol{k}-\boldsymbol{k}'\right)\,. \tag{1.105}$$

Here P_{Φ_p} is the power spectrum of the primordial fluctuations. A common parameterisation of the primordial power spectrum is the following

$$P_{\Phi_p}(k) = 2\pi^2 A_{\mathrm{s}} k^{-3}\left(\frac{k}{k_{\mathrm{p}}}\right)^{n_{\mathrm{s}}-1}\,. \tag{1.106}$$

In the above equation A_{s} is the scalar amplitude of the fluctuations, n_{s} is the *scalar spectral index* and k_{p} is the pivot scale, which has to be fixed conventionally. The difference $n_{\mathrm{s}} - 1$ measures the deviation from scale invariance of the power spectrum. Here scale invariance means that the dimensionless power spectrum $\Delta(k)$

$$\Delta(k) = \frac{k^3 P(k)}{2\pi^2}\,, \tag{1.107}$$

does not depend on the scale k. Indeed it is clear from the above definition and from Eq. (1.106) that if $P(k) \propto k^{-3}$ then $\Delta(k)$ does not depend on k.

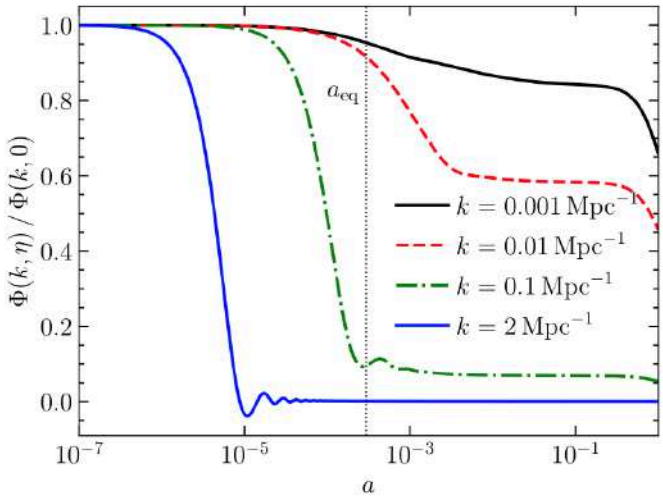

Figure 1.2: Linear evolution of Φ for modes of different scale in the fiducial ΛCDM cosmology. The solutions have been obtained using CLASS and are normalised to the value of the potential at early times. Source: [6]

1.3.4 Linear structure formation

In this section I briefly describe the linear structure evolution at late times, with the aim to write down the commonly adopted parameterisation for the matter power spectrum. The power spectrum $P(k)$ is defined as the two-point correlation function in Fourier space

$$\left\langle \delta(\boldsymbol{k})\delta(\boldsymbol{k}')\right\rangle \equiv (2\pi)^3 P_{\delta\delta}(k)\delta^{(3)}(\boldsymbol{k}-\boldsymbol{k}')\,, \tag{1.108}$$

where $\delta_{\rm m} = \delta_c + \delta_b$ is the total cold matter density contrast. The aim here is to obtain a form of the power spectrum as a function of the scale k and the scale factor a (or the redshift z) as time variable. In the linear theory the different scales k evolve independently, i.e. the evolutionary differential equations for different k are decoupled. The linear evolution of the perturbations breaks naturally into three stages: super-horizon evolution, horizon crossing, and sub-horizon evolution, corresponding to three different regimes in which the equations can be solved with specific approximations. The term "horizon" here refers to the comoving horizon defined in Sec. 1.2, which is essentially the conformal time η. Early on, all of the modes are outside the horizon ($k\eta \ll 1$) since η is small and the wavelengths of all the modes are bigger, then all modes undergo the super-horizon evolution. At intermediate times the wavelengths of some modes enter the horizon and the universe evolves from radiation domination ($a \ll a_{\rm eq}$) to matter domination ($a \gg a_{\rm eq}$). At late times all the modes have entered the horizon and therefore they evolve identically again. The different behavior according to the scale k of the modes is shown Fig. 1.2. The small scale modes of Φ enter the horizon during radiation domination, exhibiting a decaying behaviour. Large scale modes instead enter the horizon well after $a_{\rm eq}$, during the matter domination epoch, and they remain constant. A moderate decay characterises the intermediate scales, which enter the horizon around the matter-radiation equality epoch. This guides us to split the dependency on k and a of the potential Φ into two

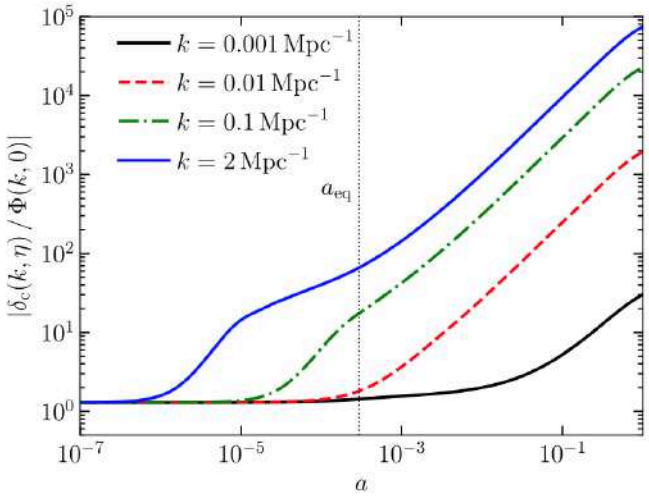

Figure 1.3: Linear evolution of the dark matter overdensity δ_c, normalised to the early times value. We see that small scales start to grow earlier, since they enter the horizon early and therefore the gravitational collapse start early. Large scale overdensities start to grow only at late times instead. At late times the growth is identical for all scales. Image taken from [6].

factors, which are the transfer function $T(k)$ and the growth factor $D(a)$

$$\Phi(\boldsymbol{k}, a) = \frac{3}{5}\Phi_p(\boldsymbol{k})T(k)\frac{D(a)}{a}\,. \tag{1.109}$$

The evolution of the dark matter overdensity is depicted in Fig. 1.3 and it can be determined from the evolution of Φ, using the late time no-radiation limit of Eq. (1.95)

$$k^2\Phi(\boldsymbol{k}, a) = 4\pi G\rho_{\mathrm{m}}(a)a^2\delta_{\mathrm{m}}(\boldsymbol{k}, a) \quad (a > a_{\mathrm{late}}\,, k \gg aH)\,. \tag{1.110}$$

Inserting the zero order Friedman equation yields an expression for the matter overdensity

$$\delta_{\mathrm{m}}(\boldsymbol{k}, a) = \frac{2}{5}\frac{k^2}{\Omega_{\mathrm{m},0}H_0^2}\Phi_p(\boldsymbol{k})T(k)D(a)\,. \tag{1.111}$$

The transfer function contains all the information about early Universe interactions, and it is conventionally set to $T(k) = 1$ for large scales, which enter the horizon at late times where no other interaction takes place but the gravitational one. The growth factor encodes the late time growth of structures, and it is proportional to the scale factor $D(a) \propto a$ during matter domination. The fact that $\delta_{\mathrm{m}} \propto D(a)$ explains where the name "growth" comes from, since it encodes the way the overdensity δ_{m} grows with the scale factor. Inserting Eq. (1.111) and Eq. (1.106) in Eq. (1.108) finally yields a parameterisation of the linear power spectrum of the matter density fluctuations

$$P^{\mathrm{L}}_{\delta\delta}(k, a) = \frac{8\pi^2}{25}\frac{A_{\mathrm{s}}}{\Omega^2_{\mathrm{m},0}}D^2(a)T^2(k)\frac{k^{n_{\mathrm{s}}}}{H_0^4 k_{\mathrm{p}}^{n_{\mathrm{s}}-1}}\,. \tag{1.112}$$

Usually the scalar amplitude A_{s} is traded in favor of the σ_8 parameter, defined as the standard deviations of the cold matter perturbations evolved in linear theory up to present time at the

scale of $8\,\mathrm{h}^{-1}\,\mathrm{Mpc}$

$$\sigma_8^2 = \int \frac{\mathrm{d}k}{k} \frac{k^3 P_{\mathrm{L}}(k)}{2\pi^2} \left|W_{\mathrm{TH}}\left(kR_8\right)\right|^2 \tag{1.113}$$

where $W_{\mathrm{TH}}(x) = 3(\sin x - x\cos x)/x^3$ is the top-hat filter in Fourier space. In the linear perturbation theory A_{s} is linearly proportional to the square of σ_8, but the latter is preferred in galaxy surveys since it is more directly measurable. An evolution equation for the growth factor $D(a)$ can be derived by considering the late time limit of the generalised Poisson equation Eq. (1.95)

$$\frac{d^2\delta_{\mathrm{m}}}{da^2} + \frac{d\ln\left(a^3 H\right)}{da}\frac{d\delta_{\mathrm{m}}}{da} - \frac{3\Omega_{\mathrm{m},0}H_0^2}{2a^5H^2}\delta_{\mathrm{m}} = 0\,. \tag{1.114}$$

This equation in general has to be solved numerically; anyway in the late Universe, where only matter and dark energy are relevant, an integral expression can be obtained

$$D(a) = \frac{5\Omega_{\mathrm{m},0}}{2}\frac{H(a)}{H_0}\int_0^a \frac{da'}{\left(a'H\left(a'\right)/H_0\right)^3}\,. \tag{1.115}$$

If the the dark energy is not a cosmological constant, the above is not a solution of Eq. (1.114), but there exists an empirical fitting for the logarithmic growth rate $f(a)$

$$f(a) \equiv \frac{d\ln D(a)}{d\ln a} \simeq \left[\Omega_{\mathrm{m},0}(a)\right]^{\gamma}\,, \tag{1.116}$$

where $\Omega_{\mathrm{m},0}(a) \equiv 8\pi G\rho_{\mathrm{m}}(a)/3H^2(a)$ is the time-dependent matter density parameter. The exponent γ is named growth parameter and its value in ΛCDM is $\gamma \simeq 0.55$ [18]. In some modified gravity (MG) theories γ has a different value instead, therefore by measuring it is possible to test these MG theories gravity against ΛCDM. Regarding the transfer function instead there exist analytical models which are valid only on certain regimes. However nowadays these models are not as useful as before, since the advent of fast and accurate codes for numerical integration of the Einstein-Boltzmann equations like CAMB [19] and CLASS [20]. These codes have reached an agreement of 0.01% [20] and they can accurately compute the power spectrum even introducing non-linearity effects, using recipes fine-tuned on cosmological N-body simulations, such as HALOFIT [21], which is based on the halo model.

1.4 Observations

After having written down all the relevant equations to study the expansion history of the Universe, we are now able to contextualise the most important cosmological observations. In this section I am reporting about

- Cosmic Microwave Background (CMB);
- SuperNovae Ia;
- Galaxy and weak lensing surveys;
- Weak lensing surveys.

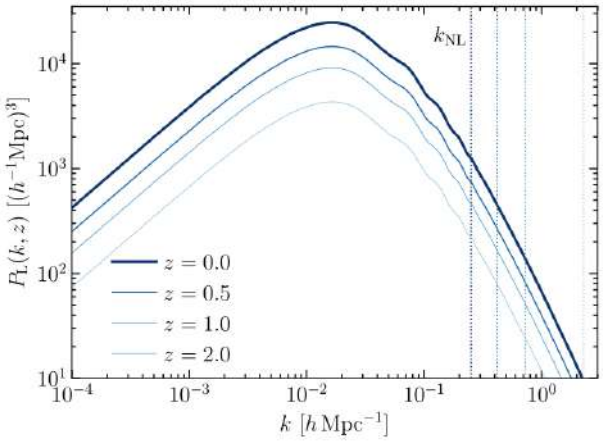

Figure 1.4: The linear matter power spectrum as function of k for different values of the redshift z in the fiducial ΛCDM cosmology. The vertical lines with the $k_{\rm NL}$ label indicate the scale at which the non-linearities arise. Source: [6].

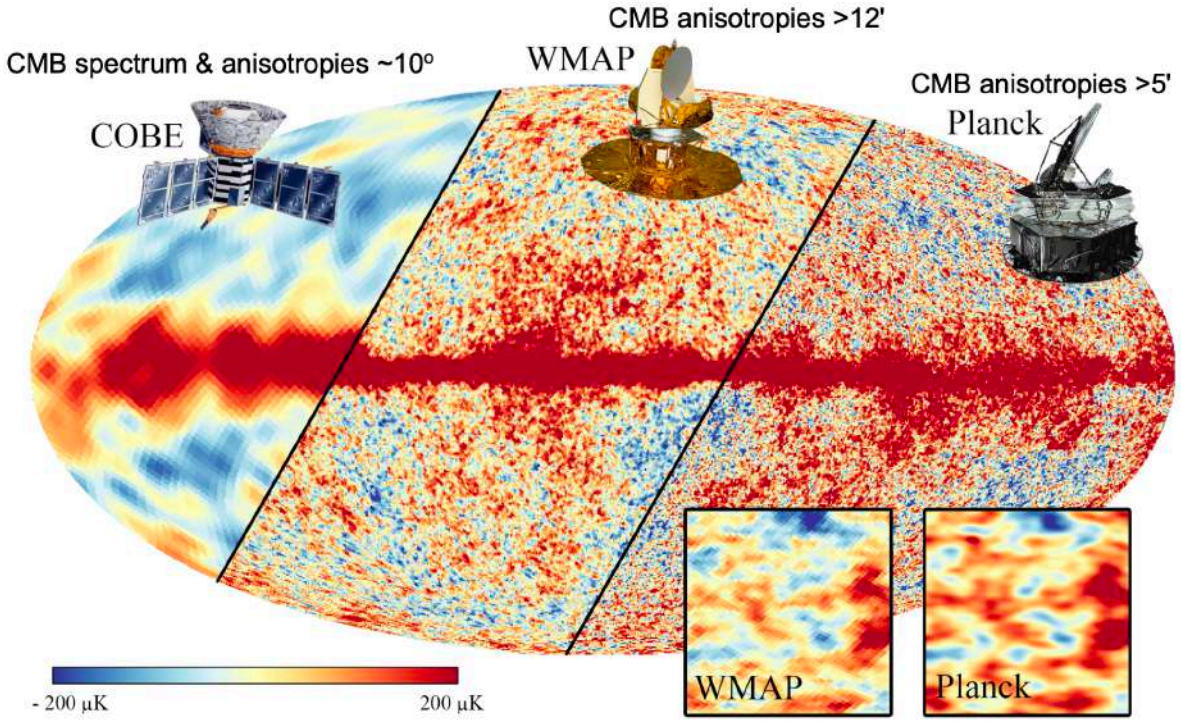

Figure 1.5: Heatmap representing the evolution of the CMB measurements throughout the years. Image taken from the plenary talk of Nabila Aghanim at the "28th Texas Symposium on Relativistic Astrophysics". Link to the presentation: `https://indico.cern.ch/event/336103/contributions/786561/`

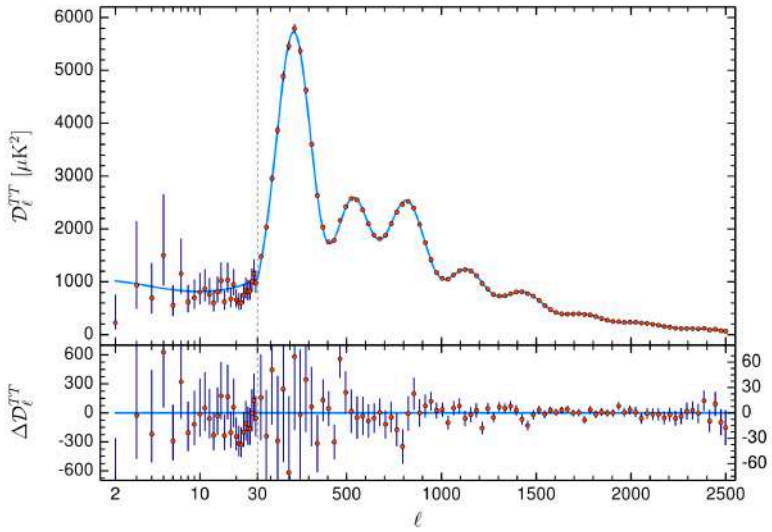

Figure 1.6: *Upper panel*: Anisotropies of the CMB as measured by the Planck satellite (red points); the function plotted is defined as $\mathcal{D}_\ell \equiv \ell(\ell+1)C_\ell T_0^2/2\pi$. The blue line shows the best-fit prediction from the ΛCDM model, with the initial conditions provided by inflation. *Lower panel*: Difference between data and best-fit model. Image taken from [6]

1.4.1 CMB

As illustrated in Sec. 1.3.1, in the early Universe the photons were tightly coupled with baryons through Compton scattering, forming a hot and dense plasma almost at thermodynamical equilibrium. During the expansion the Universe got cooler and the interactions between photons and baryons were progressively driven away from equilibrium. Photons decoupled completely at $z^* \simeq 1100$, approximately 380 000 years after the Big Bang, at the epoch of *recombination*. This name comes from the fact that at that time the free protons and electrons started to form neutral atoms, and the temperature of the photon gas was not sufficient to ionise these atoms. Therefore after recombination the Universe became transparent to the photons, which started to free-stream in the empty space, giving rise to what we call now the Cosmic Microwave Background (CMB). The CMB photons conserved anyway the imprints of the interactions taking place in the primeval plasma, becoming therefore a snapshot of the early Universe.

The first measurement of the CMB was done by Penzias and Wilson in 1965 [22]. With the 20-foot horn-reflector antenna at the Crawford Hill Laboratory they detected an excess temperature of about 3.5 K which was isotropic, unpolarised and free from seasonal variations. As also confirmed in [23] Penzias and Wilson were observing a black body spectrum with that temperature, which was the first evidence of the Cosmic Microwave Background Radiation.

However the CMB is not perfectly isotropic, as it was first observed by the COBE [24] experiment in 1992, which detected fractional temperature fluctuations Θ, modelled as in Eq. (1.54), with an order of magnitude of $\sim 10^{-5}$. Most of the physical information we can extract from the CMB is contained in its angular temperature fluctuations. The cause for these anisotropies resides in the density fluctuations which were once present in the primeval baryon-photon plasma.

These inhomogeneities got imprinted in the distribution of the photons, and turned into the anisotropies we observe today as the photons free-streamed from the Last Scattering Surface. Figure 1.5 shows the evolution of the measurements of the CMB anisotropies in the years, showing three generations of instruments with increasing precision: COBE (1992), WMAP (2003) [25], and Planck [4].

The most recent measurements of the CMB anisotropies are from Planck, and reached an angular precision of 5' [26]. The physical quantity used for quantifying the anisotropies is the fractional temperature fluctuation Θ of the CMB photons gas defined in Eq. (1.54). Since this quantity is observed as a function of the angles on the sky sphere, it is convenient to decompose it on the spherical harmonics basis

$$\Theta(\theta,\phi) \equiv \frac{\delta T(\theta,\phi)}{\bar{T}} = \sum_{\ell} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\theta,\phi) \tag{1.117}$$

The summary statistics employed by the data analysis are the angular power spectra, usually called the $C(\ell)$'s

$$C(\ell) = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2 \tag{1.118}$$

A plot of the CMB power spectrum is shown in Fig. 1.6; the agreement between the data points taken by Planck and the best-fit model is quite impressive. From the CMB spectrum, measurements of the cosmological parameters have been obtained [4], for example the Hubble constant, which was found to be

$$H_0 = (67.4 \pm 0.5)\,\mathrm{km\,s^{-1}\,Mpc^{-1}} \tag{1.119}$$

1.4.2 Supernovae Ia

A supernova is a powerful and luminous stellar explosion, occurring at the end of the life cycle of massive stars. Supernova may be some times brighter than its host galaxy, and this feature makes them easy to spot. Moreover supernovae (SNe) can be used as standard candles [27], since the mechanism generating a supernova is quite universal. This means that all SNe shine with the absolute luminosity, and therefore they can be used as distance indicators.

Supernovae are found by repeatedly observing the same patches of the sky, and then performing pixel-by-pixel subtractions of the CCD images obtained at different epochs. A particular class of SNe has been employed as standard candle, the SuperNovae Ia. The supernovae are classified by their spectra, and SNe Ia has the peculiar characteristic of having Silicon (Si) emission lines, not present in other types of supernovae. The main observable for SNe Ia measurements is the so called distance modulus μ, defined by

$$\mu \equiv 5 \log_{10}\left(\frac{d_L(z)}{10\,\mathrm{pc}}\right) , \tag{1.120}$$

with d_L being the luminosity distance defined by Eq. (1.44). From SNe Ia Riess et al. [1] in 1998 obtained the first evidence for the Universe accelerated expansion and the existence of a cosmological constant. With the years the measurements of SNe got more refined, and always Riess et al. [28] in 2019 measured the following value for the Hubble constant

$$H_0 = (74.03 \pm 1.42)\,\mathrm{km\,s^{-1}\,Mpc^{-1}}\,. \tag{1.121}$$

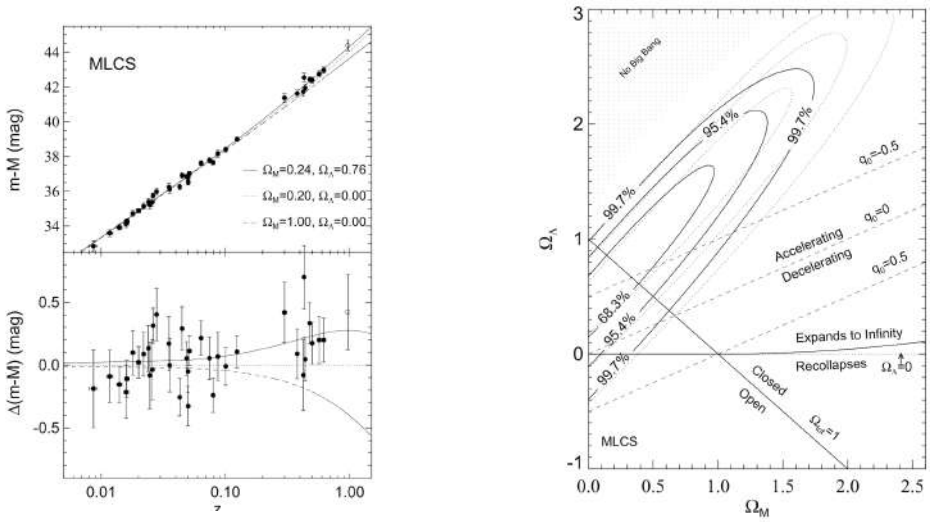

Figure 1.7: Hubble diagram and contour plots from Riess et al [1].

This results in a 4.4σ tension with the Planck measurement Eq. (1.119) from CMB anisotropies. This is one of the open problems of Modern Cosmology: the authors of [28] claim that their measurement Eq. (1.121) provides a stronger evidence for Physics beyond the Standard ΛCDM model. However there is not a common agreement in the community about the reasons of this tension, and therefore to possible solutions to it.

1.4.3 Galaxy redshift surveys

In a galaxy redshift survey the angular positions and the redshift of galaxies are measured, thus allowing to build a 3D map of the distribution of the galaxies themselves. The angular positions are quite easy to measure, while the redshifts are more complicated. The galaxies spectra are employed to measure the wavelength of prominent known emission lines; then by comparing these wavelength to the rest frame ones in the laboratory an estimate of the redshift is obtained.

Galaxy surveys can be classified mainly in two categories: spectroscopic and photometric. In a spectroscopic survey the spectra of the observed galaxies are measured with an high wavelength sampling, thus achieving high precision in the measurement of the redshift. However this advantage comes with a cost: the procedure for obtaining the spectrum of a galaxy is expensive, and this limits the number of sources whose redshifts are measured. In photometric surveys the redshift determination is based on imaging instead of spectroscopy: the photometry of galaxies is taken in various optical filters with a broad wavelength range. These images are then combined to construct a low resolution spectrum of the galaxy, and comparing it to a known template spectrum the redshift is estimated. This allows to produce a larger galaxy catalogue, since photometry is easier to obtain than spectroscopy, but at the cost of a bigger

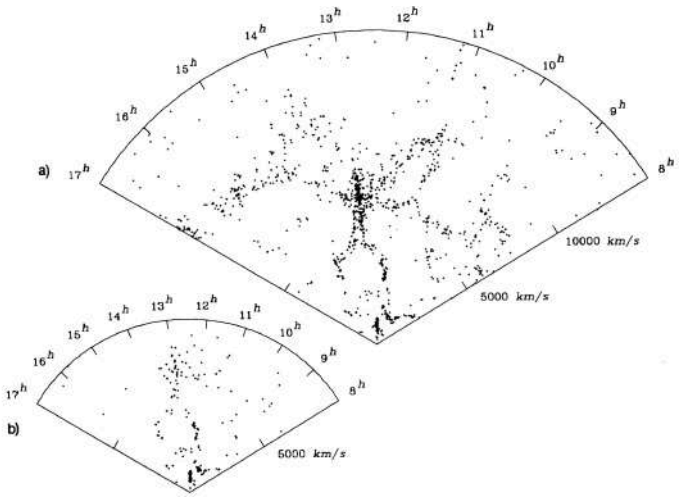

Figure 1.8: First strip of the CfA survey, showing a supercluster of galaxies. The shape of the cluster resembles a standing man, and for this reason it is referred also as the "CfA Stick Man". Source: [29].

error on the redshift of the single source.

The first galaxy survey was the Center for Astrophysics Redshift Survey [30] (CfA), which took place between 1977 and 1982, measuring the redshifts of about 2000 galaxies. Fig. 1.8 displays the results of the survey: the plot clearly shows that galaxies are not distributed randomly but along filaments forming pattern.

The Sloan Digital Sky Survey [11] (SDSS) confirmed this pattern, as we see from one of its sky maps in Fig. 1.1. Moreover with the Extended Baryon Oscillations Spectroscopic Survey [31] (eBOSS), the SDSS allowed to probe the so called Baryonic Acoustic Oscillations (BAO), which are an imprint of the early Universe left in the large scale structures we see today.

As discussed in Sec. 1.3.1, early on the Universe was an hot and dense plasma of baryons and photons, tightly coupled to each other via Compton scattering. The radiation pressure was big enough to counter gravitational collapse, and the alternating dominance of pressure and gravity was giving rise to acoustic oscillations in the primeval plasma. After the photons decoupled at recombination epoch, these sound waves froze in the baryons fluid, and their characteristic length scale became imprinted in the distribution of galaxies we observe today. SDSS measured the galaxy radial correlation function and found a peak at a comoving distance of about $100\,\mathrm{h}^{-1}\,\mathrm{Mpc}$, as reported in Fig. 1.9. SDSS was the first to observe that there is an excess of galaxies separated by that distance. This distance can be predicted from theory to be the length scale of the acoustic waves of the primordial baryon-photons fluid, and since it is a comoving scale it is usually referred as a *standard ruler* for Cosmology.

An upcoming redshift galaxy survey is *Euclid*, a space-based mission which will measure the photometric redshift of about 1.5 billion galaxies in the range $0.001 < z < 2.5$, and the spectroscopic redshift of about 50 million galaxies in the range $0.9 < z < 1.8$. The next chapter is devoted to a detailed description of the *Euclid* mission, which is the context of this thesis work.

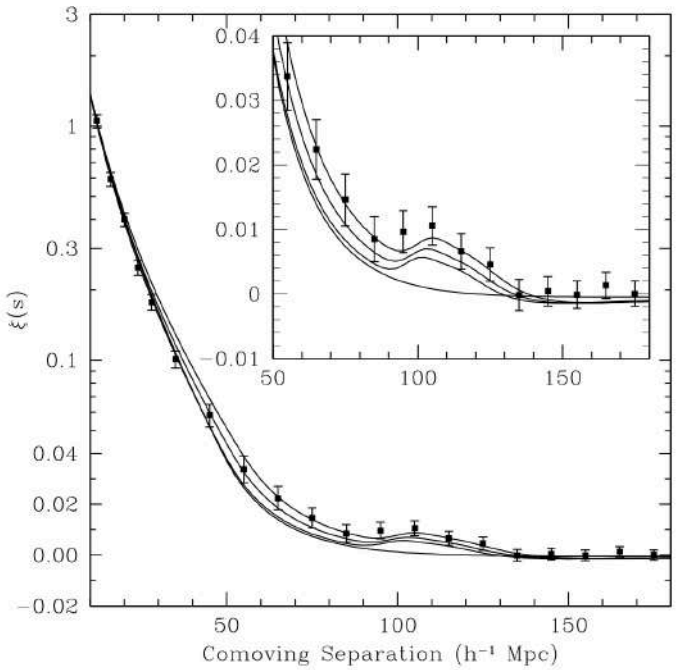

Figure 1.9: Large-scale redshift-space correlation function of the SDSS luminous red galaxies sample. The plots clearly shows that there is a peak around $100\,\mathrm{h}^{-1}\,\mathrm{Mpc}$, that has been interpreted as the BAO length scale. [32]

1.4.4 Weak lensing surveys

One of the successful predictions of General Relativity is *gravitational lensing*, i.e. the deflection of light rays due to the propagation in an inhomogeneous gravitational field, and it is encoded in the geodesic equation Eq. (1.49) for photons. Qualitatively we can distinguish between two regimes of gravitational lensing: *strong* lensing and *weak* lensing.

In the strong lensing regime the lens is a single massive object, and its gravitational field is intense enough to produce multiple images of the same source. By weak lensing instead we mean the deflection of light rays emitted by distant sources, caused by the tidal gravitational field produced by cosmological matter density fluctuations [33]. This effect is observed in the distortion of the images of light-emitting galaxies, and this distortion is directly related to the matter distribution projected along the line of sight. These image deformations can be modelled as a statistical field named *cosmic shear*, which is sensitive to the distribution of all matter, dark and baryonic, since lensing is a purely gravitational effect. Therefore, cosmological parameters can be measured by exploiting this probe, in a so called weak lensing survey, as it has been made in [34].

There have been various surveys since the first evidences [35–37] for weak lensing from large-scale structures. The Canada–France–Hawaii Telescope Lensing Survey (CFHTLens) [38] survey started in 2003 and lasted for 6 years, observing 154 square degrees of sky, creating a catalogue of about 1.5 million galaxies with median redshift of 0.75. From analysis of CFHT data, constraints [39] on cosmological parameters have been obtained, showing a tension [40] with Planck in the $\Omega_{\mathrm{m},0} - \sigma_8$ plane. This tension has been shown to be relatively robust to

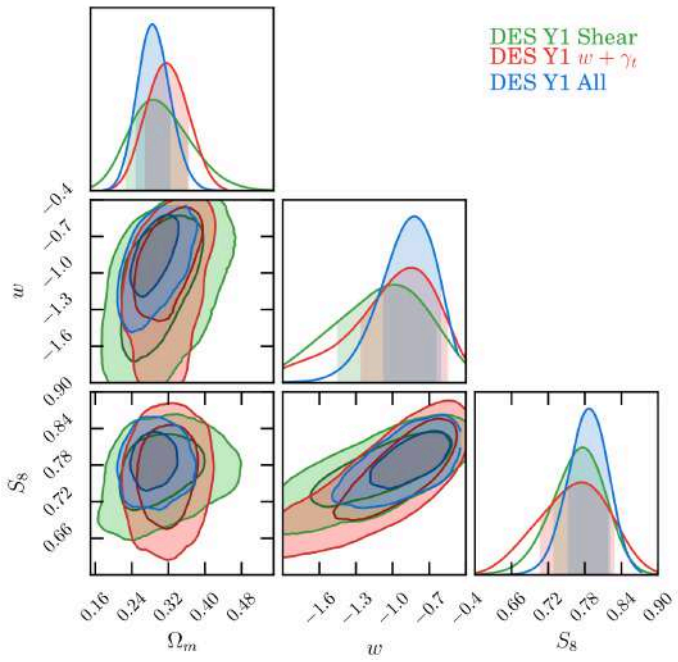

Figure 1.10: Constraints on the three cosmological parameters σ_8, $\Omega_{\mathrm{m},0}$, and w in a wCDM cosmology from the analysis of DES Y1 [41] data, after marginalising over nuisance parameters. With γ_t it is denoted the angular cross-correlation between weak lensing and galaxy clustering.

the choice of non-linear modelling, extension to the intrinsic alignment model, and inclusion of baryons.

The Dark Energy Survey (DES) started in 2013 and ended in 2019, imaging 5000 square degrees of the southern sky, using an extremely sensitive 570-Megapixel digital camera DECam, mounted on the Blanco 4-meter telescope at Cerro Tololo Inter-American Observatory. The DES collaboration performed a combined analysis of galaxy clustering and weak lensing, considering also their cross-correlation, extracting from the data tighter constraints [41] on cosmological parameters than considering the probes as independent, as it is shown in Fig. 1.10.

The Euclid mission will also perform a weak lensing survey measuring the shapes of about 1.5 billion galaxies in the visible range. In the context of my thesis, I performed a Fisher forecast of the Euclid sensitivity to the cosmological parameters, by combining weak lensing, photometric galaxy clustering, and spectroscopic galaxy clustering (see Chapter 4).

1.5 The Standard Model of Cosmology

The currently accepted cosmological model in the community is the so called Standard Model of Cosmology, which is also referred to as the ΛCDM model. This model essentially describes a flat Universe that is dominated today by non-baryonic cold dark matter (CDM) and a cosmological constant (Λ), with initial perturbations generated by inflation in the very early universe. In the previous section we have seen how the ΛCDM theoretical predictions can be tested against cosmological observations. It is remarkable how accurately the CMB spectrum

Fig. 1.6 is predicted by the ΛCDM, and how the need of a cosmological constant emerges to explain the SNe Ia Fig. 1.7 observations, from which it came the first evidence of the late Universe accelerated expansion. The cosmological parameters of ΛCDM are the following

- H_0: the Hubble parameter at present time, also referred as the Hubble constant;
- $\Omega_{\mathrm{m},0}$: the cold matter density parameter at present time;
- $\Omega_{\mathrm{b},0}$: the cold baryonic matter density parameter at present time;
- σ_8: the standard deviations of the cold matter perturbations evolved in linear theory up to present time at the scale of $8\,\mathrm{h}^{-1}\,\mathrm{Mpc}$;
- n_s: the scalar spectral index, quantifying the scaling exponent of the primordial perturbations.

$H_0\,[\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}]$	$\Omega_{\mathrm{m},0}$	$\Omega_{\mathrm{b},0}h^2$	σ_8	n_s
(67.36 ± 0.54)	0.3153 ± 0.0073	0.02237 ± 0.00015	0.8111 ± 0.0060	0.9649 ± 0.0042

Table 1.1: Base ΛCDM cosmological parameters as measured by Planck [4]. The uncertainties are the 1-σ marginalised constraints, i.e. at 68% CL. For the baryon density $\Omega_{\mathrm{b},0}$ parameter Planck reports the constraints only on the product $\omega_{\mathrm{b},0} \equiv \Omega_{\mathrm{b},0}h^2$.

The values of these parameters as recently measured by Planck [4] are reported in Tab. 1.1.

Despite its success, the ΛCDM model has some problems: first of all we do not know the physical nature of dark energy and dark matter, which together constitute the 95% of the Universe energy content. Concerning dark energy, in the ΛCDM model it is theoretically implemented as a cosmological constant (Λ), which is characterised by an energy density which is constant in space and time.

Due to this characteristic, there has been the attempt to interpret Λ as the contribution of the quantum zero-point energy, or *vacuum* energy, a concept which is comprised in Quantum Field Theory. However when calculations are done to verify this hypothesis, the expected vacuum energy density predicted by Quantum Field Theory turns out to be 120 orders of magnitude [42] greater than the value coming from observations. This is known as the *cosmological constant problem*, and it has not been solved yet.

There have been various attempts to model dark energy other than Λ, e.g. dynamical dark energy models. A commonly employed dark energy model is the Chevallier-Polanski-Linder (CPL) [43, 44] parameterisation

$$w_{\mathrm{DE}}(z) = w_0 + w_a \frac{z}{1+z}\,, \tag{1.122}$$

to which it corresponds the following analytical redshift evolution of the dark energy density

$$\rho_{\mathrm{DE}}(z) = \rho_{\mathrm{DE},0}(1+z)^{3(1+w_0+w_a)} \exp\left[-3w_a \frac{z}{1+z}\right]. \tag{1.123}$$

The extension of the ΛCDM model in which dark energy follows the CPL parameterisation is usually referred as the w_0w_aCDM model, since it is a minimal extension in which the only

difference with respect to ΛCDM is that dark energy is parameterised by Eq. (1.122). In a w_0w_aCDM cosmology the redshift evolution of the Hubble parameter can be written in closed form as

$$H(z) = H_0\sqrt{\Omega_{\mathrm{m},0}(1+z)^3 + \Omega_{\mathrm{DE},0}(1+z)^{3(1+w_0+w_a)}\exp\left(-3w_a\frac{z}{1+z}\right) + \Omega_{k,0}(1+z)^2}\,. \tag{1.124}$$

Another issue of ΛCDM is the tension on the value of the Hubble constant H_0; in the previous section it has been explained that early Universe probes, such as the CMB spectrum measured by Planck, give a value Eq. (1.119) which results in a 4.4σ tension with the one Eq. (1.121) coming from late Universe probes, such as the SNe Ia. There is not yet a common agreement about the reason for this tension: it can be due to unaccounted observational systematics or to new physics beyond ΛCDM. On the one hand the SNe Ia must be calibrated in order to get a reliable measurement of H_0, since it is degenerated with the absolute magnitude of SNe. On the other hand the Planck analysis of the CMB spectrum is model dependent; this means that different results may come out using a different cosmological model.

In the next chapters I will extensively talk about the *Euclid* mission, which could shed new light on these aspects, and test the ΛCDM concordance model with more accuracy.

Chapter 2

The Euclid Mission

Euclid is a space-based mission of the European Space Agency (ESA) aimed to investigate the nature of dark energy and dark matter by observing the formation of large scale structures over cosmological timescales [45, 46]. The mission is scheduled for launch by the end of 2022, and the spacecraft will travel to the L_2 Sun-Earth lagrangian point, where it will observe the extra-galactic sky for approximately 6 years.

The *Euclid* survey will cover $15\,000\,\mathrm{deg}^2$, which is about 36% of the sky, using two main probes for constraining cosmological parameters: Weak Lensing (WL) and Galaxy Clustering (GC). The former requires measuring the shapes and photometric redshifts of the galaxies, the latter is based on the measurement of the three-dimensional distribution of galaxies through their spectroscopic redshifts. For achieving these tasks *Euclid* will mount two instruments on board: a visible imager (VIS) and a near-infrared photometer and spectrometer (NISP). With VIS the shape deformations of 1.5 billion galaxies will be measured, adopting a 6×6 array of 4k CCD detectors sensitive in the visible range. NISP will instead be equipped with 4×4 HgCdTe detectors sensitive in the near-infrared range ($900 - 2000\,\mathrm{nm}$), and it will operate in two modes: photometric and spectroscopic. In photometric mode, it will employ three broad-band filters to estimate the photometric redshifts of the galaxies imaged by VIS, in order to maximise the information gain coming from the weak lensing survey using tomography, as we will detail later. In spectroscopic mode, a *grism*, a dispersing element made by a prism with a diffraction grating, will be used to observe the spectra of 50 million galaxies. Combining the two main probes, *Euclid* will reach unprecedented precision in the measurement of fundamental cosmological parameters.

In this chapter I am giving a description of the *Euclid* mission. The chapter is structured as follows. In Sec. 2.1 I present the scientific goals of the mission. In Sec. 2.2 I report about the structure of the satellite, giving some details about the VIS and NISP instruments. In Sec. 2.3 I describe the Euclid survey, detailing the observational strategy of a telescope pointing, also outlining the scientific requirements for the main probes. Finally in Sec. 2.4 I give a description of the management and the organisation of the mission.

2.1 Scientific objectives

Euclid has been conceived to measure the Universe expansion history and the growth of large-scale structures with high precision and accuracy. The main scientific goals of the mission

can be summarised in four topics, each being the parameterisation of a extension of the Concordance ΛCDM model described in Sec. 1.5. Here I briefly describe each of them.

Dark Energy The equation of state of dark energy in ΛCDM is characterised by $w = -1$ at all redshifts, i.e. a cosmological constant. As explained in Sec. 1.5 a widely adopted way to parameterise deviations from a cosmological constant is the CPL Eq. (1.122) parameterisation. *Euclid* aims to measure w_0 and w_a with a precision of ~ 0.01 and ~ 0.1 respectively, corresponding to a *Figure-of-Merit* [1] (FoM) of about 400, thus giving unprecedented constraints for distinguishing the cosmological constant from a more complex dynamical dark energy model.

Modified gravity Another possibility for explaining the cosmic acceleration is a breakdown of Einstein's General Relativity (GR) on cosmological scales, which would be accompanied by the need of theories alternative to GR, commonly referred as Modified Gravity (MG) theories. In some of these models, deviations from GR are captured by a change in the growth of structures in the Universe, quantified by the parameter γ introduced in Eq. (1.116), which has a value of ~ 0.55 in ΛCDM. *Euclid* will measure this parameter with a 1σ precision of 0.02, accurately testing the ΛCDM model, since any detected deviation from $\gamma = 0.55$ would represent new physics beyond the concordance model.

Massive neutrinos Oscillation experiments have shown that neutrinos do have mass, determining the mass-splitting between the energy eigenstates. However directly measuring the absolute mass scale of neutrinos is extremely difficult for the current generation of particle experiments such as Katrin [14].

As pointed out in Sec. 1.3.1 from a cosmological point of view massive neutrinos can be considered warm dark matter. They are practically subject only to gravity and they are light; for this reason they can free-stream out of dense regions. This yields to a small suppression in the small scale matter power spectrum, whose scale is directly related Eq. (1.89) to the sum $\sum m_\nu$ of the neutrino masses, and therefore to the absolute neutrino mass scale. This effect will observed by *Euclid* with a $\Delta m_\nu < 0.03\,\mathrm{eV}$ accuracy, which is sufficient to distinguish between normal and inverted mass hierarchy if the total mass $\sum m_\nu$ will turn out to be $< 0.1\,\mathrm{eV}$.

Initial conditions In the concordance model the primordial spectrum of fluctuations resulting from inflation is assumed to be gaussian and nearly scale invariant. Primordial non-gaussianities are predicted also by the simplest inflationary models, and are quantified by the f_{NL} parameter. The scalar spectral index n_{s}, defined in Sec. 1.5, quantifies instead the scale-dependency of the primordial power spectrum, and scale invariance means $n_{\mathrm{s}} = 1$. *Euclid* will measure f_{NL} with an accuracy of $\sim 2\%$, which will be an improvement of the accuracy of $\sim 5\%$ reached by Planck [4].

2.1.1 Main probes

Euclid will achieve the above scientific goals described above with a galaxy survey which is on an unprecedented regime in terms of size. It will map the dark matter distribution through

[1] The Figure of Merit of a dark energy experiment is defined by $1/(\Delta\tilde{w}_0 \times \Delta\tilde{w}_a)$, where $\tilde{w}_0$ and $\tilde{w}_a$ are linear combinations of w_0 and w_a chosen such that they are uncorrelated [46].

Euclid Mission Summary

Main Scientific Objectives

Understand the nature of Dark Energy and Dark Matter by:

- Reach a dark energy $FoM > 400$ using only weak lensing and galaxy clustering; this roughly corresponds to 1 sigma errors on w_p and w_a of 0.02 and 0.1, respectively.
- Measure γ, the exponent of the growth factor, with a 1 sigma precision of < 0.02, sufficient to distinguish General Relativity and a wide range of modified-gravity theories
- Test the Cold Dark Matter paradigm for hierarchical structure formation, and measure the sum of the neutrino masses with a 1 sigma precision better than 0.03eV.
- Constrain n_s, the spectral index of primordial power spectrum, to percent accuracy when combined with Planck, and to probe inflation models by measuring the non-Gaussianity of initial conditions parameterised by f_{NL} to a 1 sigma precision of ~2.

SURVEYS

	Area (deg2)	Description
Wide Survey	15,000 (required) 20,000 (goal)	Step and stare with 4 dither pointings per step.
Deep Survey	40	In at least 2 patches of > 10 deg^2 2 magnitudes deeper than wide survey

PAYLOAD

Telescope	1.2 m Korsch, 3 mirror anastigmat, f=24.5 m				
Instrument	VIS	NISP			
Field-of-View	0.787×0.709 deg^2	0.763×0.722 deg^2			
Capability	Visual Imaging	NIR Imaging Photometry			NIR Spectroscopy
Wavelength range	550– 900 nm	Y (920-1146nm),	J (1146-1372 nm)	H (1372-2000nm)	1100-2000 nm
Sensitivity	24.5 mag 10σ extended source	24 mag 5σ point source	24 mag 5σ point source	24 mag 5σ point source	3 10^{-16} erg cm-2 s-1 3.5σ unresolved line flux
Detector Technology	36 arrays 4k×4k CCD	16 arrays 2k×2k NIR sensitive HgCdTe detectors			
Pixel Size Spectral resolution	0.1 arcsec	0.3 arcsec			0.3 arcsec R=250

SPACECRAFT

Launcher	Soyuz ST-2.1 B from Kourou
Orbit	Large Sun-Earth Lagrange point 2 (SEL2), free insertion orbit
Pointing	25 mas relative pointing error over one dither duration 30 arcsec absolute pointing error
Observation mode	Step and stare, 4 dither frames per field, VIS and NISP common FoV = 0.54 deg^2
Lifetime	7 years
Operations	4 hours per day contact, more than one ground station to cope with seasonal visibility variations;
Communications	maximum science data rate of 850 Gbit/day downlink in K band (26GHz), steerable HGA

Budgets and Performance

	Mass (kg)		*Nominal Power (W)*	
industry	TAS	Astrium	TAS	Astrium
Payload Module	897	696	410	496
Service Module	786	835	647	692
Propellant	148	232		
Adapter mass/ Harness and PDCU losses power	70	90	65	108
Total (including margin)	**2160**		**1368**	**1690**

Figure 2.1: Summary of the *Euclid* mission from the "Euclid Definition Study Report" [46].

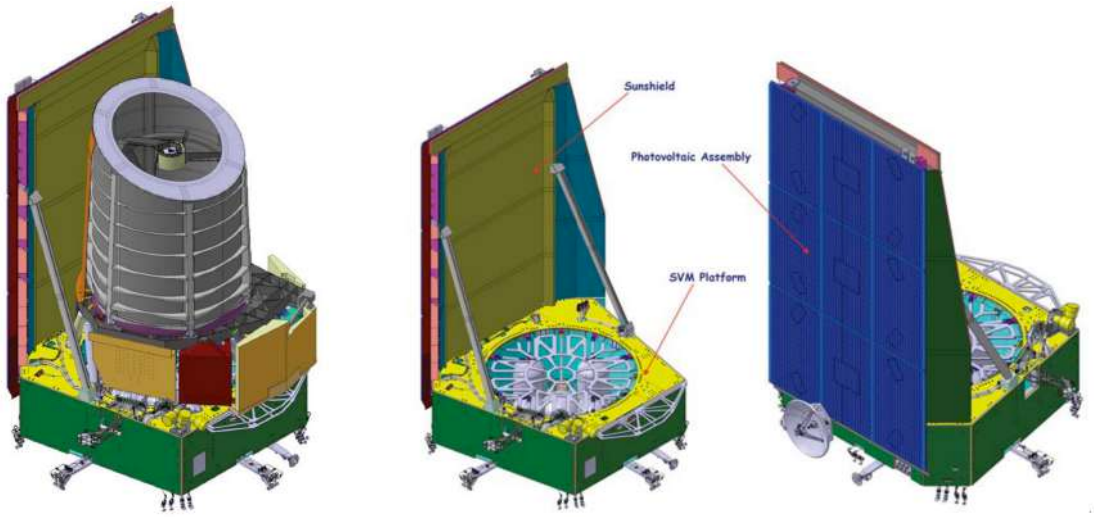

Figure 2.2: Overview of the Euclid spacecraft (left) and Service Module (right) [45].

weak gravitational lensing by imaging with VIS about 1.5 billion galaxies, also providing estimates of their photometric redshifts through near-infrared photometry, with a nominal precision [46] of $\sigma_z^{\rm ph}/(1+z) \leq 0.05$. Even if weak lensing is a cumulative effect, depending on all the matter along the line of sight, the knowledge of the photometric redshifts of the sources allows to partially recover the three-dimensional information via the so called weak lensing *tomography* [47]. This technique consists in slicing the redshift distribution of the sources in a given number of redshift bins, thus partially reintroducing some information of the structure along the line of sight. In [47] it is shown that even a crude tomography using two or three redshift bins is sufficient to improve, by up to an order of magnitude, the measurements of cosmological parameters coming from the weak lensing signal.

At the same time *Euclid* will perform a spectroscopic redshift survey with NISP, providing the spectroscopy of 50 million galaxies in the redshift range $0.9 < z < 1.8$, with a required accuracy of $\sigma_z^{\rm sp}/(1+z) \leq 0.001$ [46]. The spectroscopic redshifts will be measured via the detection of the Hα emission line, which has a wavelength of $\lambda \simeq 656.3\,$nm. The desired redshift range $0.9-1.8$ corresponds to have a spectroscopic instrument sensitive in the range $1100-2000\,$nm, which is near-infrared (NIR). This justifies the need for a space mission, since NIR light is completely absorbed by the atmosphere, so it would be impossible to probe this range of the electromagnetic spectrum with an Earth-based experiment.

2.2 Spacecraft

The *Euclid* spacecraft can be subdivided in three main parts: a Service Module (SVM), a Payload Module (PLM), including the telescope, and the Scientific Instruments. The overall spacecraft envelope, which is depicted in the left panel of Fig. 2.2, has dimensions of 4.8 m height and 3.74 m diameter. The SVM, shown in the right panel of Fig. 2.2, comprises the spacecraft subsystems supporting the payload operation, hosts the warm electronics of the payload, and provides structural interfaces to the PLM and the launch vehicle. The Sunshield,

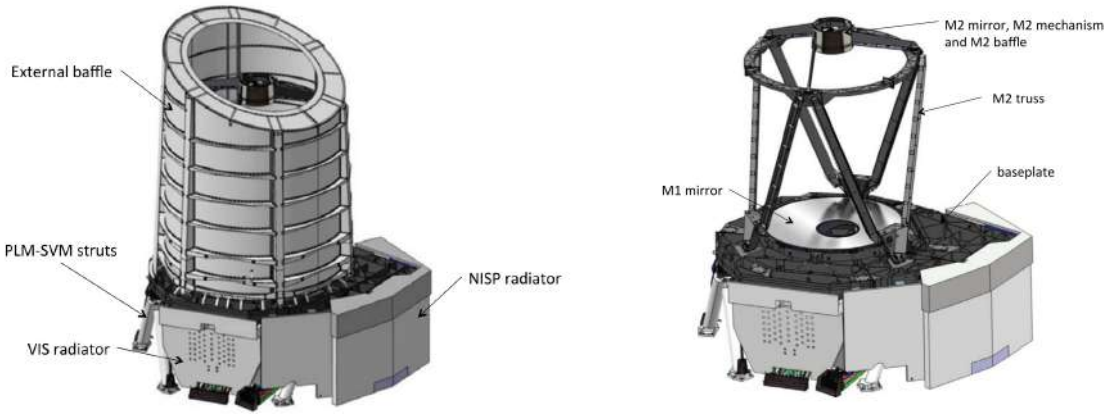

Figure 2.3: Overview of the Euclid PLM with the external baffle (left) and without it (right) [45].

part of the SVM, protects the PLM from illumination by the sun and supports the photovoltaic assembly supplying electrical power to the spacecraft. The PLM is designed around a three mirrors anastigmatic Korsch Silicon Carbide (SiC) telescope feeding the two instruments, VIS and NISP, as it is shown schematically in Fig. 2.4. The separation of the light between the two instruments is performed by using a dichroic plate, which is located at exit pupil of the telescope. Moreover the PLM provides mechanical and thermal interfaces to the instruments (radiating areas and heating lines), and it is divided in two cavities, which are separated by the base plate:

- The front cavity including the telescope primary and secondary mirrors as well as the M2 refocusing mechanism and the associated support structure. This cavity is thermally insulated by a baffle (see Fig. 2.3).
- The instrument cavity including the telescope folding mirrors, the tertiary mirror, the dichroic splitter, the two instruments (VIS and NISP), the shutter and calibration source for the VIS channel.

The PLM mechanical architecture is based on a common SiC baseplate which supports on one side telescope M1 and M2 mirrors and on other side the other optics and the two instruments. This architecture is well adapted to the selected thermal architecture with telescope and instrument cavities both passively controlled at neighbour temperatures. The optical accommodation on the baseplate consists in implementing two folding mirrors (FoM1 and FoM2) at the entrance of the instrument cavity (between M2 and M3) to fold the optical beam in the plane of the baseplate. A third folding mirror (FoM3) allows having the VIS instrument close to an efficient radiative area.

2.2.1 VIS instrument

VIS [49] is the Visible Imager of *Euclid*: the core task of VIS is to enable Weak Lensing measurements, which are possible by measuring in an accurate way the galaxy shapes. VIS is

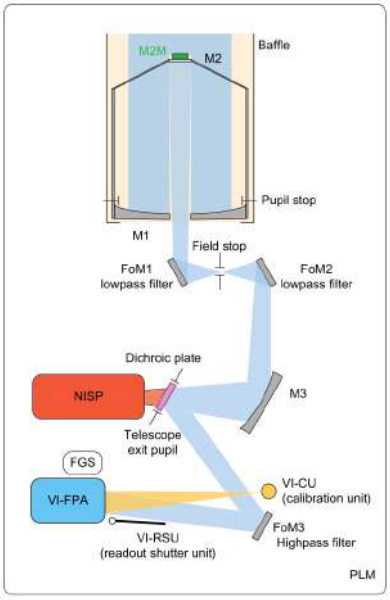

Figure 2.4: Functional view of the Euclid Payload Module (PLM) [45].

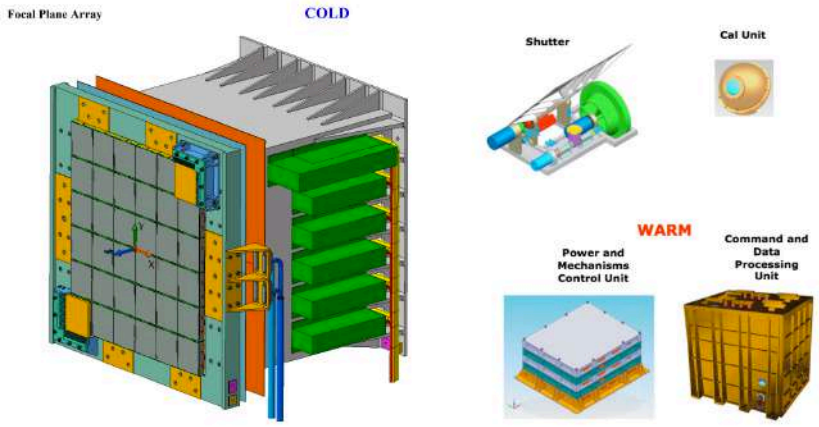

Figure 2.5: Units comprising VIS: Power and Mechanisms Control Unit and Command and Data Processing Unit in the bottom right are in the Service Module, while the Focal Plane Array, the Shutter and the Calibration Unit are in the Payload Module. Source [48].

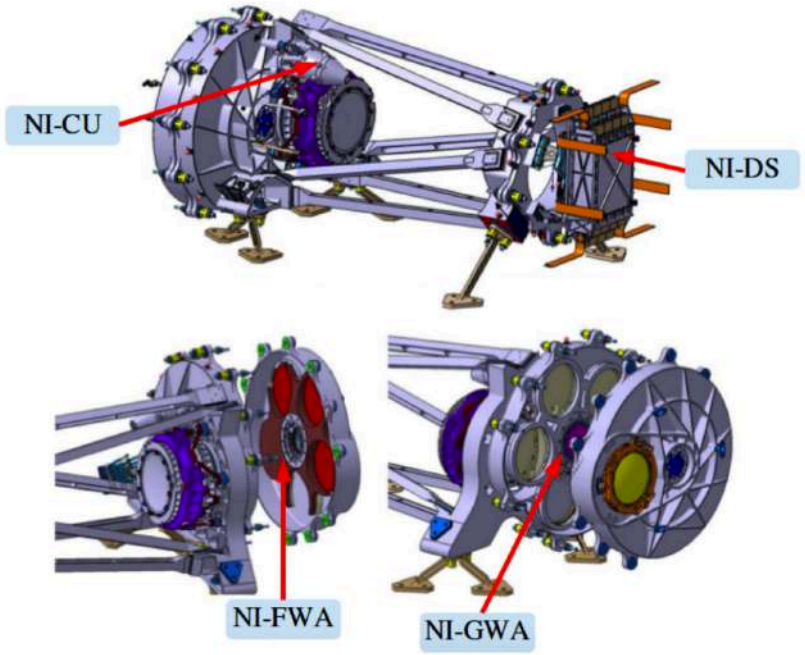

Figure 2.6: NISP view from [50].

sensitive in the visible range $550 - 900\,\text{nm}$, and as shown in Fig. 2.5 it is composed by various parts:

- The Focal Plane Array (VI-FPA), which consists in a matrix of 6×6 CCDs made by 4k×4k pixels. Each pixel is $12\,\mu\text{m}$ square with an angular scale of 0.1 arcsec, and this corresponds to a total field of view of $0.5\,\text{deg}^2$.
- The calibration unit (VI-CU), which is designed to allow flat fields of the visible channel. This structure encloses a 12-LEDs panel illuminating a diffusing panel inside an integrating sphere.
- The shutter mechanism, which is closed right after an observation ends, in order to prevent direct light from falling onto the CCDs during the data processing phase.
- Two electronics units, one for the control and processing of instrument data, the other for the control of the mechanisms and the power allocated to the instrument; the central data processing unit receives data from CCDs and compresses the resulting large image (24k×24k) in about 250 seconds; the power and mechanisms control unit switches the shutter and the calibration unit on/off.

2.2.2 NISP instrument

NISP [50] is the Near-Infrared Spectrometer and Photometer, and it is designed to carry out photometry and slitless spectroscopy in the near infrared. Its main components are:

- a filter wheel (NI-FWA), holding the three filters (Y, J and H bands) for the photometric mode and performing the switching function between these. Each filter is 10 mm thick

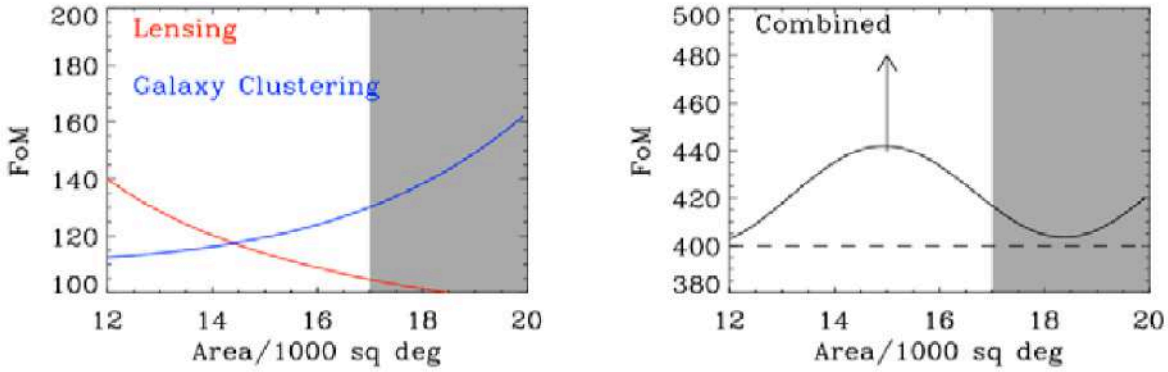

Figure 2.7: Plots of Figure-of-Merit for Galaxy Clustering and Weak Lensing (left panel), and the combination (right panel) as a function of area, fixing the total survey time to 5 years. There exists a sweet spot for the FoM resulting from the combinations of the main probes, which is around 15 000 deg^2. Source [46]

and 120 mm in diameter. The wheel is configured with five slots in order to integrate in addition an open slot and a closed one.

- a grism wheel (NI-GWA), hosting the three grisms employed for dispersing light in spectroscopic mode. It positions into the optical beam two kinds of grisms with different passband coatings. The first type (Blue) transmits between 1100 nm and 1450 nm, while the other (Red) is transparent between 1450 nm and 2000 nm. The 4 grisms are mounted in two orientations: blue 0 deg, red 0 deg, blue 90 deg and red 90 deg.

- a focal plane (NI-DS), equipped with an array of 4×4 HgCdTe detectors made by 2048×2048 pixels sensitive in the near-infrared range $900 - 2000$ nm. The size of the focal plane is 16 cm × 16 cm, with a field of view (FoV) of 0.5 deg^2, therefore each pixel has an angular scale of 0.3 arcsec;

- the warm electronics unit (NI-WE), which is located in the Service Module and allows the synchronous acquisition of the focal plane detectors, as well as in-flight data pre-processing;

- a calibration unit (NI-CU), located close to the filter wheel.

2.3 The Euclid Survey

As pointed out in Sec. 2.1 *Euclid* aims to reach a FoM > 400 from the combination of weak lensing and galaxy clustering. Weak lensing alone would favour smaller and deeper survey, while the constraints from galaxy clustering improve when wider and shallower areas are considered, as it is shown in the left panel of Fig. 2.7. When the combinations of the two probes is considered, a sweet spot around 15 000 deg^2 is found, as shown in the right panel of Fig. 2.7.

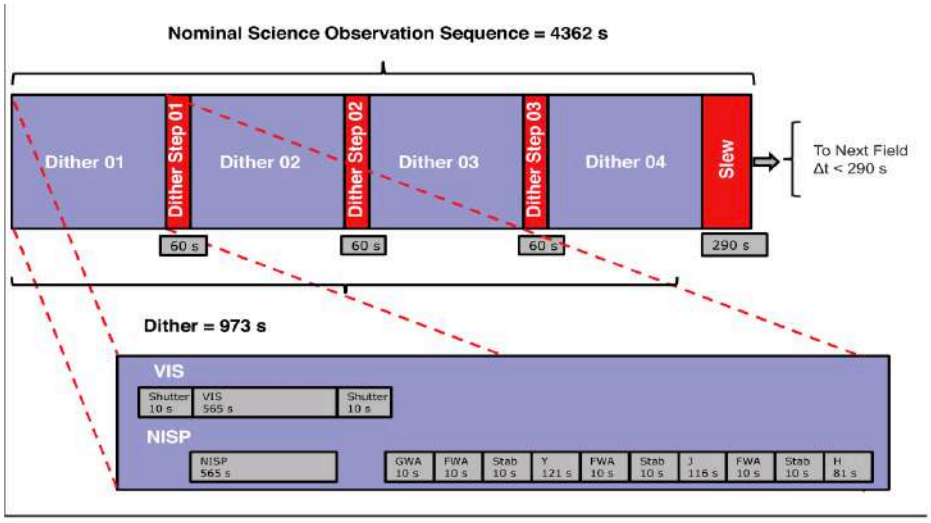

Figure 2.8: Schematic timeline of an observation sequence of one field. Each frame (blue) starts with a simultaneous exposure of VIS and NISP in spectroscopic mode, followed by the three NISP photometric exposures. The dither-to-dither slews and the field-to-field slew are marked in red. Source: [46].

2.3.1 Observational sequence

The elementary observation sequence of the survey is named *field*. As it is illustrated in Fig. 2.8 each field is composed of four frames of the 0.54 deg^2 common field of view, observed with a dither step in-between. During each frame VIS and NISP carry out exposures of the sky simultaneously; at the end of the last frame a slew to the next field is performed. As it can be seen from Fig. 2.8, in the first 610 s VIS and NISP take data for weak lensing and spectroscopy respectively. Then NISP photometric imaging follows, with integration times 88 s, 90 s and 54 s for filters Y, J, and H respectively. During the NIR photometry, VIS closes its shutter in order to avoid disturbances coming from the movement of the filter wheel. There are multiple reasons for which this dithering strategy has been designed. First, the focal planes of VIS and NISP have gaps between the active surfaces of the detectors, and since the focal planes of the two instruments have different layout these gaps do not coincide. As Fig. 2.9 shows the dither movements are designed to cover the gaps and appropriately sample the field of view. Second, this observation strategy also mitigates the impact of cosmetic defects and cosmic rays on science data. Finally in the case of NISP spectroscopy, the four frames are used to measure the spectra with two grism rotations to minimise source confusion due to overlapping spectra.

2.3.2 Weak lensing and the photometric survey

The gravitational lensing signal from large-scale structure represents a $\sim 1\%$ change in a galaxy ellipticity so, to extract a significant signal, a large number of galaxies are required. To be useful for weak lensing, a galaxy must be sufficiently distant, to maximise the lensing effect which is cumulative along the line of sight, and resolved, to allow for shape measurements. In order to reach a FoM > 400 as detailed in the previous paragraph, the required number of galaxies to be surveyed turns out to be at least 30 per arcmin^2 [46]. Moreover, in order to

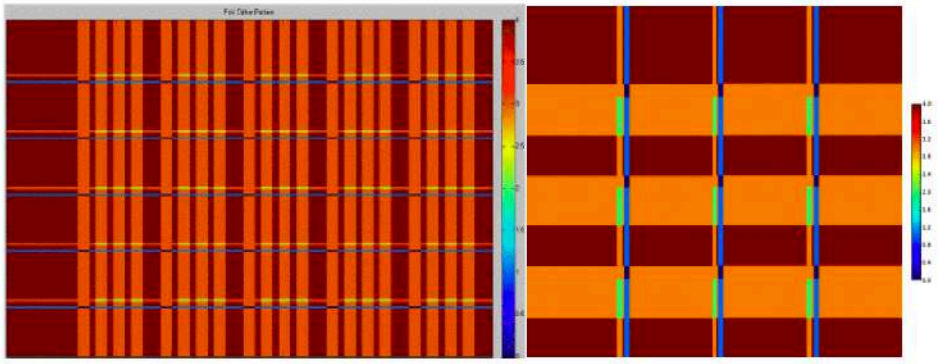

Figure 2.9: Field coverage provided by the dither pattern for VIS (left) and NISP (right). Deep red colours indicate the fraction of sky within a field covered with four frames, orange with three frames, yellow-green with two frames and blue with only one. Source: [46].

fully exploit the weak lensing signal, the tomographic technique will be employed. For this reason, estimates of galaxy redshifts are needed for the majority of sources used in the weak lensing analysis.

The photometric redshift measurement relies on finding the best galaxy spectral template match to the broad-band colours, and the accuracy depends on the number of filters and the signal-to-nose ratio of the observations. In particular the required accuracy for the estimate of the photometric redshifts is that the standard deviation with respect to the true (spectroscopic) redshifts is $\sigma_z^{\rm ph}/(1+z) \leq 0.05$ [46].

The NIR photometry provided by *Euclid* data alone is not sufficient to achieve the required photometric redshift accuracy and precision, thus additional ground-based data are required. The survey area has to be imaged from the ground using the griz filters, which cover the full wavelength range $420-930$ nm and have overlap of less than 10%. Finally, the mean redshift in each tomographic bin to be used is required to be known better than $\sigma(\langle z \rangle) < 0.002(1+z)$, and must be unbiased. Therefore, the set of templates used to estimate the photometric redshifts needs to be representative of the true galaxy spectra. These templates can be obtained from ground-based spectroscopic observations of a sufficiently large (at least 1×10^5) number of galaxies, with a fraction of incorrect redshifts sufficiently small (less than 1×10^{-4}).

2.3.3 Spectroscopic survey

The constraints coming from the galaxy clustering (GC) improve with the number of galaxies whose redshifts are observed, as it could be expected. The theoretical limit to GC performance is the cosmic variance, which is due to the limited volume of the survey. In order to meet the FoM requirements from the combination of galaxy clustering with weak lensing, *Euclid* will employ slitless spectroscopy to obtain the spectroscopic redshifts for a minimum of 3500 galaxies per deg^2, which yields a sample of 52.6×10^6 galaxies over the total sky area (15 000 deg^2) covered. The requirement about the accuracy of the spectroscopic redshift measurement is $\sigma^{\rm sp}(z)/(1+z) < 0.001$.

The main spectral feature that will be used for the measurement of the redshift is the Hα line, which comes from the transition between the $n = 3$ and $n = 2$ energy levels of

Req. ID	Parameter	Requirement	Goal
WL.1-5	Redshifts error ($\sigma(z)/(1+z)$)	≤ 0.05	≤ 0.03
WL.1-6	Catastrophic failures	10%	5%
WL.1-7	Error in mean redshift in bin	<0.002	
WL.2.1-17	NIR wavelength range	920 to ≥1600nm	
WL.2.1-18	NIR number of filters:	≥3	
WL.2.1-19	NIR PSF size:	EE50 and EE80 Y: (<0.30", <0.62") J: (<0.30", <0.63") H: (<0.33", <0.70")	
WL.2.1-20	NIR Pixel scale:	0.3±0.03 arcsec	
WL.2.1-21	Relative Photometric Accuracy	<1.5%	

Figure 2.10: The top level photometric requirements from weak lensing for the *Euclid* survey. Source [46].

hydrogen and has a wavelength of $\lambda \simeq 656.3\,\mathrm{nm}$. In particular the science requirement for the number density of the galaxies in the spectroscopic sample translates into demanding for the average effective Hα flux limit from a 1 arcsec diameter source to be lower than or equal to $3 \times 10^{-16}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$ at 1600 nm. The flux limit[2] is defined as the line flux for which the signal to noise ratio is > 3.5. Because of the number and the size of the detectable spectra in the field of view, a large number of spectra will be contaminated (or confused) by spectra from other galaxies. This contamination leads to failures in the redshift measurement, which means that not all detected spectra will allow to extract a reliable measurement of the redshift. The two main parameters for determining the quality of the resulting galaxy catalogue are *completeness* and *purity*. The completeness C is defined as the ratio between the number N_{meas} of the spectra with a detected Hα flux which is above the specified limit and the total number N_{tot} of detected galaxies:

$$C = \frac{N_{\mathrm{meas}}}{N_{\mathrm{tot}}}\,. \tag{2.1}$$

The estimate of N_{tot} comes from the current knowledge of the spatial density and luminosity function [3] of Hα emitting galaxies. The scientific requirement for the completeness of the *Euclid* spectroscopic catalogue is to reach a completeness higher than 45% [46]. The purity of the sample is instead defined as the fraction of correctly measured redshifts among the measured ones. Let N_{corr} the number of correct redshifts, then the purity p is given by the ratio

$$p = \frac{N_{\mathrm{corr}}}{N_{\mathrm{meas}}}\,. \tag{2.2}$$

A too low purity can reduce the ability of the survey to detect the BAO signal: the threshold set for this parameter is $p > 80\%$. To achieve the specified completeness and purity, NISP will have to obtain multiple exposures of the same field of view at different dispersion directions. This will be done to reduce the confusion originated from contaminated spectra, which is intrinsic to the slitless technique. This is the reason why NISP will be equipped with more

[2]The flux limit is defined to be the lower limit above which all galaxies must be detected.

[3]The luminosity function of galaxies is defined as the distribution (histogram for simplicity) of the galaxies against their luminosity.

Req. ID	Parameter	Requirement	Goal
GC.1-2	Galaxy sky density	3,500 / deg^2	5,000 / deg^2
GC.1-8	Bias of all galaxies	>1	
GC.1-9	Bias, upper quartile in redshift	>1.3	
GC.2.1-1	Flux limit	$\leq 3\times10^{-16}$ erg cm^{-2} s^{-1}	
GC.2.1-2	Completeness	>45%	
GC.2.1-3	Flux limit at all wavelengths	<120% of GC.2.1-1	

Figure 2.11: The top level requirements for galaxy clustering measurement from *Euclid* survey. Source [46].

than one grism with different orientations. Together with spectroscopy, NIR imaging in the photometric channel of NISP will be used to associate the spectra to the objects in the field of view, in order to remove ambiguities with zero orders of the spectra themselves. The NIR photometric counterpart of the field of view will also give the size the orientation of the objects, with which the best extraction aperture [4] of the spectra can be determined, and will also provide flagging of the contaminated spectra. In particular, NIR imaging in the H band to a depth of AB magnitude < 24 will be sufficient to meet the needs of spectroscopy.

2.4 Mission organisation

The *Euclid* Ground Segment is in charge of managing the mission from the ground. It is split into two blocks: the Operational Ground Segment (OGS) and the Science Ground Segment (SGS). The OGS is managed entirely by ESA, and comprises the Mission Operations Centre (MOC) and the ground stations network. The SGS instead is under the shared responsibility of ESA and the Euclid Consortium, this last being the group of researchers in charge of data processing, instrument operations, survey definition, and archiving, with the final aim to fulfill the scientific goals of the mission. The main components of the Ground Segment, schematically represented in Fig. 2.12, are:

- The Mission Operations Centre (MOC): it is located in ESOC, Darmstadt, and is in charge of all mission operations planning, execution, monitoring and control of the spacecraft and ground segment operations;

- The Ground Stations: being part of ESA's tracking network, they are under control of the MOC at ESOC. The network is composed of 3 Deep Space antennas in New Norcia (NNO), Western Australia, Cebreros (CEB), Spain, and Malargüe, Argentina, which are used during Launch and Early Operations Phase (LEOP), the commissioning phase and the routine mission. A small X-band antenna will also be available in New Norcia (NNO) for first acquisition during LEOP. Some of the 15 m ESTRACK stations in Kourou (KOU), French Guiana, and Maspalomas (MSP), will also be employed during LEOP, if still available at the time. The ground stations will provide the communication

[4] The extraction aperture of a spectrum is the image window which will be used to convert the 2D spectrum produced by the grism on the detector to the 1D spectrum needed for redshift measurement.

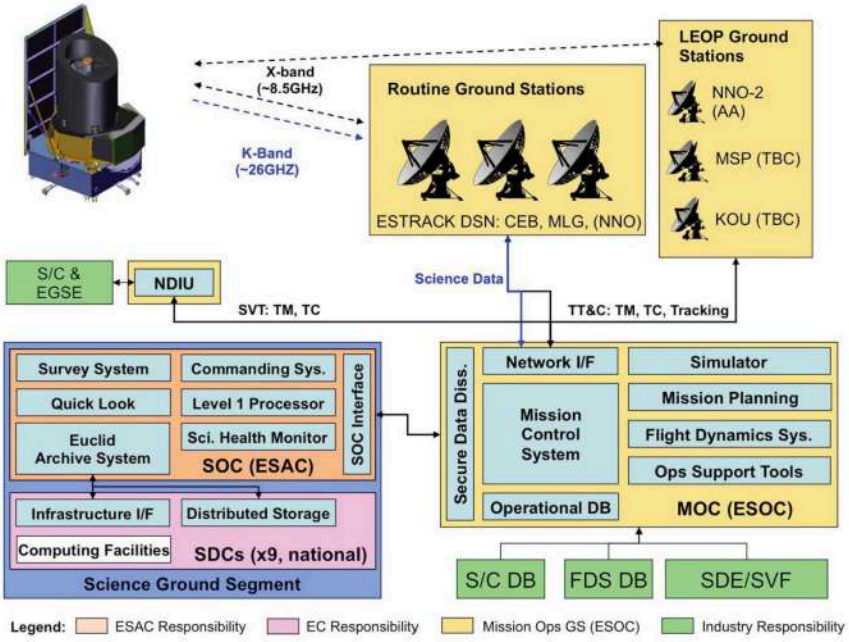

Figure 2.12: Overview of the *Euclid* Ground Segment. Source: [45].

with the spacecraft during all mission phases post-launch in X-band on up- and downlink. Two of the 35 m stations will undergo an upgrade for receiving Euclid's 26 GHz signal.

- The Science Operations Centre (SOC): it is located in ESAC, Villafranca, and has the duty to plan scientific operations, monitor the performance of the payload using spacecraft and instrument files delivered by the MOC. It will also interface with the Euclid Science Data Centres (SDC), in order to archive and distribute science data and to furnish support for the scientific analysis.

- The Communications Network, which links the various remote centres and stations to support the operational data traffic.

- The industrial prime contractor, in charge of providing the spacecraft and flight dynamics databases together with onboard software images to the MOC.

2.4.1 Science Ground Segment

The Science Ground Segment (SGS) is composed mainly by two parts: the Science Operations Center (SOC) and the Euclid Consortium. The SOC is managed by ESA, and interfaces with the MOC. On the one hand the MOC provides raw telemetry and all the necessary information to manage the mission from the scientific point of view; on the other hand the SOC provides the MOC with information related to observation planning and instruments commanding. The Euclid Consortium (EC) is the fraction of the Ground Segment (ECSGS) performing the data processing, starting from the telemetries coming from the satellites down to the data products

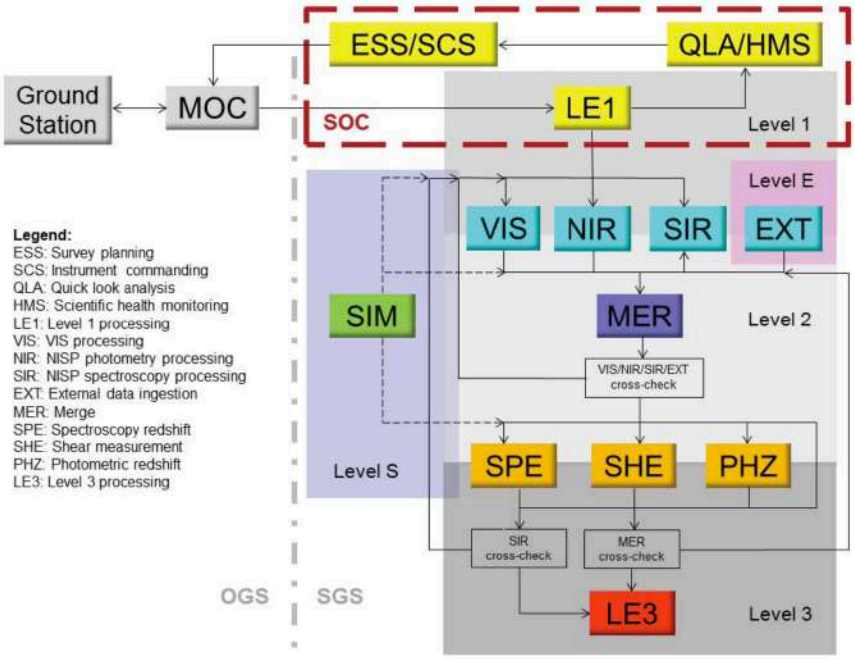

Figure 2.13: Functional decomposition of the *Euclid* Science Ground Segment (SGS). Source: [45].

of the mission. The ECSGS is composed of a number of Science Data Centres (SDCs), in charge of managing all science-driven data processing, from instrument data to simulations and science products. In the EC are also present two Instrument Operation Teams (IOTs), one for VIS and another one for NISP, whose task is to guarantee instrument maintenance and operations. The two IOTs coordinates with each other, providing thus a single interface between the ECSGS and the SOC. In Fig. 2.12 the SGS is represented by the box on the lower-left corner: the Euclid Archive System (EAS) is at its heart, since it has been conceived with the role of managing data in a centralised way for both the EC and the general scientific community.

The scientific data processing of science data can be decomposed in ten Processing Functions (PFs), representing self-contained processing units communicating through the EAS. Fig. 2.13 schematically shows the interrelations between the various PFs, which are listed in the following:

- **LE1**: it provides unscrambled telemetry data, the so-called level 1 data;
- **VIS**: it processes the visible imaging data from level 1 data to fully calibrated images;
- **NIR**: it produces the fully calibrated Near-Infrared images data, starting from LE1 data, together with source lists, for quality check purposes and to allow spectra extraction;
- **SIR**: it is in charge of extracting sources spectra and fully calibrated spectral images, starting from LE1 spectroscopic observations;

- **EXT**: it ingests in the EAS all external data needed to complement Euclid data;
- **SIM**: it produces the simulations needed to test, validate and check the quality of the whole pipeline;
- **SPE**: it measures spectroscopic redshifts from the SIR extracted spectra;
- **PHZ**: it estimates the photometric redshifts from the NIR imaging data;
- **SHE**: it measures shapes and ellipticities of galaxies from VIS imaging data;
- **MER**: it performs the merging of all information, providing stacked images and source catalogues, aggregating photometric and spectroscopic data;
- **LE3**: it computes the high-level science data products, the so called Level 3 data, starting from the fully processed shape and redshift measurements and any other possibly needed Euclid data.

Each Processing Function is supported by an Organisation Unit (OU) that design algorithms and prototypes starting from the input data of the PF, at the end comparing the results of the processing with the original requirements. Once validated by the OU, the algorithms are passed on to an SDC, which turns them in full-fledged *Euclid* pipeline elements. The SGS also contains a SGS System Team (SGS ST) common to both SOC and ECSGS, which helps the SGS to define the data processing philosophy, architecture and strategy. Among the tasks of the SGS ST there are: writing code guidelines; producing tools to support software test and integration; designing and implementing common software and the Euclid Archive System; producing common Euclid Data Models, for efficient and smooth communication between the PFs with a commonly adopted data format.

The whole *Euclid* pipeline periodically undergoes reviews and functionality checks, the so called Scientific Challenges (SCs). This mechanism probes the status of the system, verifying that it is practically feasible, scalable and able to be run smoothly in a distributed environment with consistent results. The SCs are implemented in all SDCs, and at the moment of writing this thesis the mission is going through Scientific Challenge 8 (SC8).

In my thesis work I have been a member of OU-SIM and OU-SIR. In particular I developed a code for performing specific spectroscopic simulations for the validation of SIR algorithms, starting from the official simulator provided by OU-SIM. The main contributions of this work are described in Chapter 3.

Chapter 3

Spectroscopic simulations for SIR validation

The *Euclid* survey will perform the largest spectroscopic survey to date, collecting the spectroscopic redshift of 50 million galaxies. The surveyed volume will be about 500 times greater than the one of the SDSS [11] and the redshifts measured will span the range 0.9-1.8. The instrument devoted to spectroscopy measurements is NISP, described in Sec. 2.2, which will record two-dimensional images of the spectra similar to the one shown in Fig. 3.1. The calibrated one-dimensional spectra of the galaxies are extracted from these images. The extraction is made by the SIR Processing Function, as explained in Sec. 2.4. The extracted spectra are then passed to the SPE Processing Function, which measures the redshifts from the Hα emission line.

Slitless spectra extraction is not straightforward, and a whole devoted team working on it is needed. A software element is being written and tested specifically for this task, the SIR Pipeline. The completeness Eq. (2.1) and purity Eq. (2.2) of the spectroscopic catalogue strictly depend on the accuracy of the flux calibration of the spectra reconstructed by SIR. The accuracy on the spectroscopic redshift measurement is instead strongly influenced by the wavelength calibration of the extracted spectra. In order to guarantee the scientific quality of the spectroscopic catalogue that *Euclid* will produce, the SIR Pipeline has specific requirements to fulfill. A complex and detailed validation of the software must be performed before the launch, and for this reason a wide variety of simulations is needed.

In this chapter I report the results of the work that I have done in the first part of my PhD. I developed a software pipeline to perform automated and detailed validation of the SIR Pipeline, in order to test the fulfillment of the scientific requirements above outlined. The chapter is structured as follows: Sec. 3.1 briefly introduces the slitless spectroscopy technique employed in *Euclid*. Sec. 3.2 describes the SIR Pipeline, the data reduction tool for spectra extraction. Sec. 3.3 details how simulations are done by SIM, which the official *Euclid* PF devoted to deliver simulated data to the SGS. In Sec. 3.4 I describe how the software I wrote works, detailing the advantages of its implementation for the SIR validation procedures. Finally Sec. 3.5 presents some of the test cases in which the code I developed has been successfully used for performing such validations.

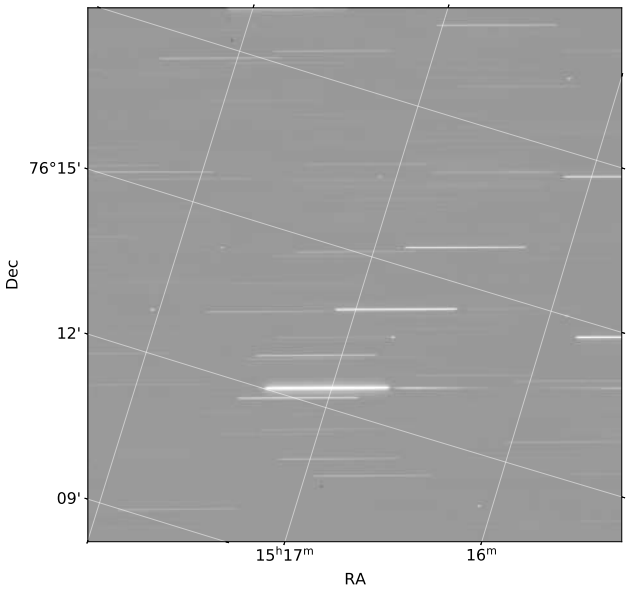

Figure 3.1: Simulated slitless image of one of the 16 NISP detector. The axes are aligned with the detector reference frame, while the sky coordinates are rotated because of the position angle of the telescope. This frame is taken from an official OU-SIM simulation, and therefore includes all instrumental effects available at the moment of production.

3.1 Slitless spectroscopy

3.1.1 Overview

The spectra of astrophysical objects can be used for multiple reasons. Detailed information about the source can be inferred from its spectrum: its chemical composition, pressure, density, and temperature. Information about the galaxy rotation speed and its radial velocity are also contained in galactic spectra. In a galaxy survey, the spectra of the galactic sources are employed for determining their redshifts. By identifying one or more spectral features, e.g. an emission line with known wavelength, the redshift of the source can be estimated. As explained in the previous chapter, in the *Euclid* survey the spectroscopic redshifts will be measured through the identification of the Hα emission line.

The spectra of the sources can be measured through instruments called spectrometers. In some surveys, the light coming from the celestial object can be observed through a long narrow aperture, named *slit*, and in this case the instrument is a slit spectrometer. The main advantage of using a slit spectrometer is the targeted nature of the observation: the slit allows one to select the source to observe, avoiding contamination from other objects. However, when the spectra of many multiple sources have to be recorded, using a slit may not be practically possible. In this case another technique, named *slitless* spectroscopy, can be adopted.

3.1.2 Features and systematics

The essence of slitless spectroscopy is that the "slits" (in analogy to slit or multi-slit spectroscopy) are defined by the position in the sky, the size, and the shapes of the sources. One can consider that an object, whose light is dispersed in a slitless image, defines its own virtual slit. The resulting spectral resolution is determined by the length of the spectrum in the dispersion direction. The size of the object in the transverse-to-dispersion direction defines instead the width of the slit.

For non-circular objects, a preferred axis, such as the major axis of a galaxy, can define a tilted virtual slit. Since the positions of the virtual slits are set by the pattern of objects on the sky, superposition of spectra may commonly occur. This *cross-contamination* happens predominantly in the dispersion direction, as the spectra are many times longer than the target along the dispersion direction. If a grism is used as the slitless dispersing element, then the zeroth and higher orders present an additional source of contamination. Grisms are usually designed in such way that the zeroth and higher orders have lower transmission than the strongest order, which often is taken to be the (positive) first order.

A further characteristic of slitless spectroscopy is the sky background level: in fact, since the maximum slit width is the detector size, the background level is many times larger than for slit spectroscopy. In *Euclid* the main background light contribution will be given by the *zodiacal* light, which is the light from the Sun scattered by interplanetary dust. Zodiacal light's intensity is maximum on the ecliptic plane, on which the Earth's orbit lies. In order to reduce as much as possible the zodiacal light, the *Euclid* survey is designed to give higher priority to regions of the sky at higher ecliptic latitudes, where this background light is less intense. Together with cross-contamination, the zodiacal light will be one of the dominant noise sources, and therefore shall be accurately modelled in order to remove it. Maps of the expected background light are available in the *Mission Database* (MDB), which is briefly described in Sec. 3.3.2.

Another systematic effect which is present in slitless spectroscopy is given by the image distortions. In general, distortions arise when the imaging scale of an optical system is not uniform across the field of view (FoV), in particular for off-axis points. The distortions can be mainly divided in two categories: *astrometric* and *spectroscopic* (or spectral). The astrometric distortions [51] cause a non-linear mapping between the coordinates of the objects on the sky and the corresponding ones on the field of view (FoV). Astrometric distortions are influenced only by the opto-mechanical layout of the instrument and can be corrected with an optical distortion model, which provides the mapping between sky coordinates and pixel coordinates on the FoV.

Spectroscopic distortions cause the deviation of the spectra traces with respect to a straight line aligned with the dispersion direction, and also a stretching or elongation of the spectra. Thus the imaged spectra are not perfectly straight, and the relation between the pixel coordinates along dispersion and the wavelength may be nonlinear. In order to correct for spectral distortions it is necessary to provide a Curvature Model and an Inverse Dispersion Solution. The Curvature Model parameterises the deviation of the spectra from straight lines, and it is usually described by a low degree (≤ 2) polynomial [51]. The Inverse Dispersion Solution provides instead the functional relation between wavelength and pixel coordinates along the dispersion direction. As for the Curvature Model, the functional relation defining the Inverse Dispersion Solution is described by a relatively low degree polynomial. The degree of

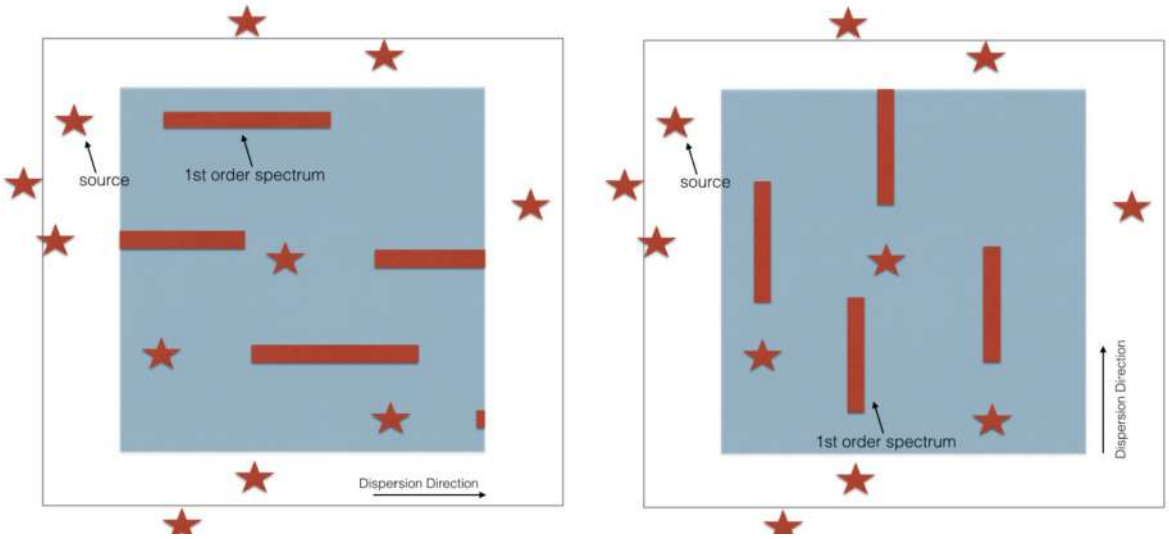

Figure 3.2: Schematic illustration of the slitless dispersed images obtained by two orthogonal grism orientations. Source [53].

the chosen polynomial depends on the grism used to obtain the data, and therefore on the observations spectral resolution.

Finally, a knowledge and modelling of the detector is needed for dealing with slitless spectroscopic data. As we outlined in Sec. 2.2 the NISP sensitive elements are semiconductor H2RG detectors produced by Teledyne, and are arrays of 2048×2048 pixels [52]. They are built using a Mercury Cadmium Telluride (HgCdTe) alloy, which is used in applications where the highest performance IR detectors are required. Essentially these detectors convert the near infrared photons coming from astrophysical sources into electron-hole pairs. The conversion rate, called *quantum efficiency* (QE) can vary across the pixel matrix. Because of the non-zero operational temperature, there exist a *dark current* in the detectors. Dark current is basically made of thermally generated electron-hole pairs, which contribute as a noise to the measured electric signal. HgCdTe detectors achieve the highest quantum efficiency and lowest dark current of any IR detector material, and for this reason are employed in the *Euclid* mission [52]. The readout noise is another source of contamination of the measured signal, and is caused by the silicon read-out arrays used to digitise the signal. All the noise sources have to be quantified by measurements on the instrument, and then implemented into the models used to simulate the data. In *Euclid*, the team directed by the *Instrument Scientist*, has the responsibility to make the connection between measurements and simulations. The results of the analysis of the detector and the optical performances are ingested into the MDB, in order to be available for simulations and science processing.

The number of galactic sources falling within a typical *Euclid* FoV is estimated to be about 10^5, so slitless spectroscopy is the only way to measure so many galaxy spectra. As said above, one of the main drawbacks of slitless spectroscopy is cross-contamination, which may reduce the number of usable sources for redshift measurement up to a factor of 50%, as mentioned in Sec. 2. The contamination issue can be solved by obtaining multiple exposures of the same field of view (FoV) with different dispersion directions: two sources whose spectra

overlap in one dispersion direction, will not overlap in the orthogonal direction. Combining the information coming from orthogonally dispersed images can reduce confusion due to the cross-contamination phenomenon.

3.2 The SIR Pipeline

In the *Euclid* mission, the SIR Processing Function (PF) is the function in charge of the reduction of the NISP spectroscopic data. The NISP Level 1 data are pre-processed by NIR, in order to flag bad or saturated pixels in a common way. Starting from these pre-processed NISP frames, SIR produces one-dimensional spectra that are fully wavelength and flux calibrated, and corrected for contamination from nearby objects. In this section I give an outline of the structure of the SIR Pipeline, the software that is being used for performing such a task, together with its scientific requirements.

3.2.1 Structure

The SIR PF is divided into two main Processing Element Blocks (PEBs): the pre-processing PEB and the extraction PEB. The former has the task of removing most detector signatures from the data. The latter is in charge of locating and extracting the spectra from the NISP spectroscopic data, and to produce fully wavelength and flux calibrated one and two-dimensional spectra. The steps composing the scientific pipeline, illustrated in Fig. 3.3, are the following:

1. coarse and fine spectra location;
2. global background estimation and subtraction;
3. extraction of 2D spectrum;
4. estimation of wavelength-dependent pixel illumination and flat fielding;
5. two-dimensional spectrum wavelength assignment;
6. two-dimensional spectrum decontamination flagging;
7. one-dimensional spectrum extraction;
8. combination of spectra between different detectors and pointing;
9. final one-dimensional spectra production.

After the pre-processing and the location phase, the 2-D spectra extraction starts. This procedure consists in cropping a small rectangular window of the whole image around each spectrum. This window has a certain *aperture*, which is the width in the direction orthogonal to dispersion. The aperture is one of the most critical parameters of the SIR pipeline, and its value is chosen by maximising the signal-to-noise ratio. For this procedure, ad hoc simulations are needed (see Sec. 3.5).

After the 2-D extraction, a re-sampling must be performed on the two-dimensional spectrum. This is necessary since all the extracted spectra have to be defined on the same common

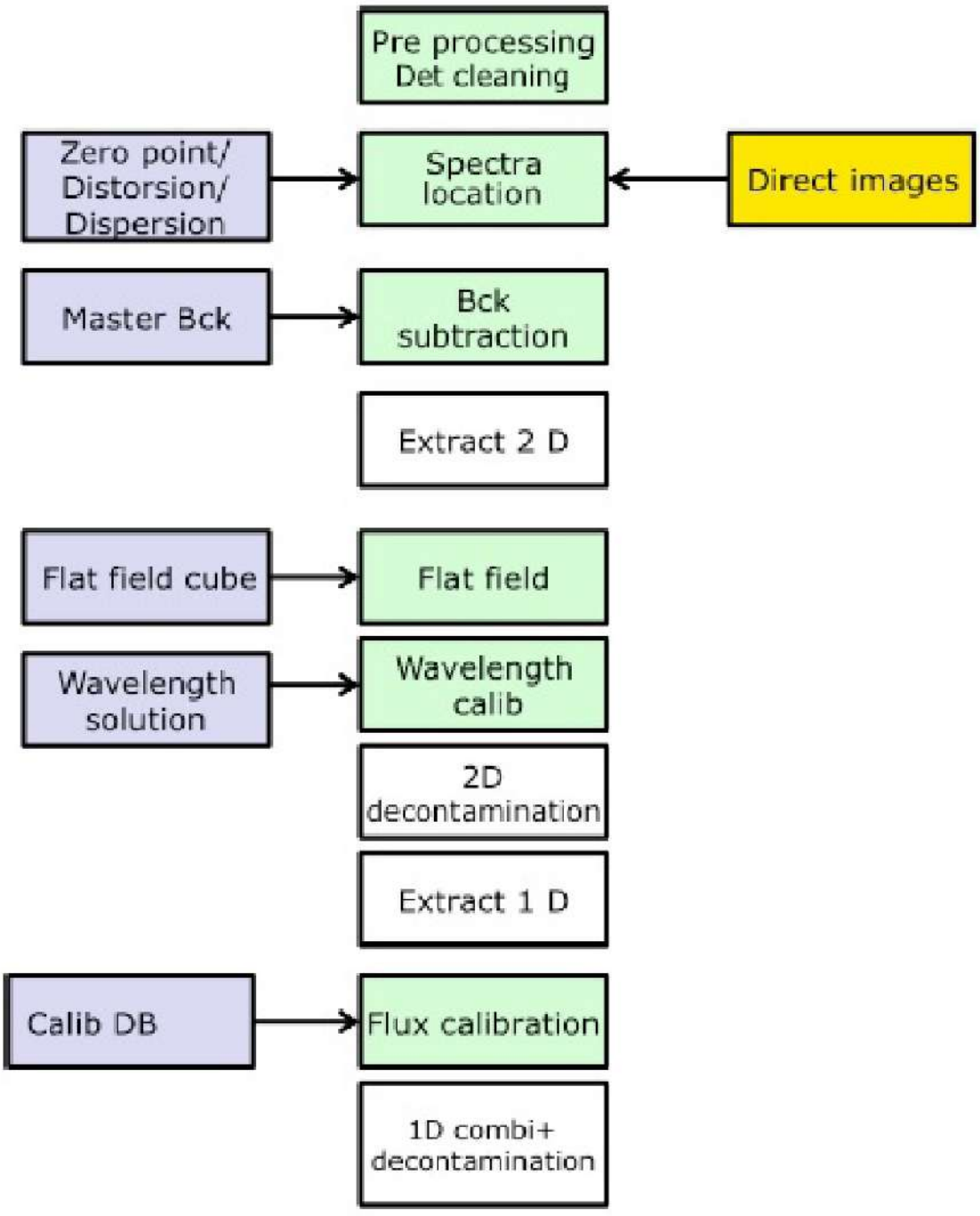

Figure 3.3: The block flow diagram of the SIR pipeline.

wavelength range. However, the relation linking pixel intervals to wavelength intervals is not uniform across the FoV, because of the spectroscopic distortions. Then each two-dimensional spectrum has to be re-sampled to translate it into the desired wavelength range.

The SIR PF tasks end with the final one-dimensional spectra production. The SPE PF successively executes the emission line fitting and the redshift determination. The final outputs are the catalogues of the objects with fitted line wavelengths and redshift. The SIR PF has some fundamental relations with other processing functions. Some of them are presented in the list below:

- from LE1, it receives raw NISP spectroscopic frames and house- keeping data;
- with NIR, it shares the task of the pre-reduction of the NISP photometric and spectroscopic frames, and the definition of the data products resulting from this pre-reduction;
- from MER, it receives the photometric target catalogue to be used for the extraction of spectra, and their de-contamination;
- from SPE, it receives requirements on the format of the extracted one-dimensional and two-dimensional spectra, which is then be delivered to SPE for redshift measurement. In validation phases, SPE returns measurements of line position and flux to SIR.

3.2.2 Scientific requirements

In parallel to the scientific pipeline, within the SIR processing function there are also a calibration pipeline and a validation pipeline. The calibration pipeline is in charge of reducing the calibration observations and of preparing the calibration data needed by the scientific pipeline. The requirements to be fulfilled by the validation pipeline are contained in the "Euclid SGS SIR Requirements Specification Document" [54]. The high-level architecture of the Pipeline originates instead from the "Euclid SGS SIR Processing Function Validation Plan" [55], which also defines the validation approach and the test cases. The validation process is divided into:

- software validation, i.e. the validation of the various processing elements that compose the SIR PF;
- data validation, i.e. the validation of the SIR PF data products.

The validation process is evolving with time, and with the availability of different sets of data to be used in the validation procedure. The scientific requirements of the SIR pipeline are inherited from the mission scientific goals. The accuracy of the spectroscopic redshift measurements, indeed, depends very strictly on the accuracy of the spectroscopic data wavelength calibration. This in turn depends on the accuracy of the dispersion solution across the NISP field of view. By requirement, the spectroscopic redshift accuracy of each detected galaxy shall be better than

$$\sigma_z < 0.001(1+z) \tag{3.1}$$

Since the requirement on redshift accuracy must be satisfied for all z, it is enough to meet the most stringent condition, i.e. $\sigma_z < 0.001$. The central wavelength of NISP spectra is $\lambda = 15 \times 10^3$ Å, then:

$$\sigma_{\Delta\lambda} < 15 \times 10^3 \text{ Å} \times 0.001 = 15 \text{ Å} \tag{3.2}$$

Since a pixel of the infrared NISP detector corresponds to 13.4 Å, we obtain a $\sigma_{\Delta\lambda} \simeq 1.1$ pixels. Such error has to be shared between wavelength zero point and dispersion solution as specified by the two following requirements contained in the requirements specification document [54]:

- "R-SIR-CAL-F-020 Wavelength Zero Point": the SIR Cal PF shall be able to determine the spectral wavelength zero point of any detected object to better than 0.63 pixels;
- "R-SIR-CAL-F-030 "Wavelength Dispersion Solution": The SIR Cal PF shall use all relevant calibration data to give a spatially varying wavelength solution across the field-of-view, accurate to a level of 0.4 pixels (rms).

Also the flux of the extracted spectra must be accurately determined by calibration procedures. Two requirements on the accuracy of the SIR PF follow from the scientific requirements of completeness and purity of the spectroscopic galaxy sample explained in Sec. 2.3.3. In particular, the SGS requirements specification document [54] specifies that:

- "R-SIR-CAL-F-060 Instrument Response Stability (Survey)": within the Wide Survey, a spatially varying solution for the relative flux calibration shall be provided. The solution shall be such that the relative flux at the flux limit of the sample, averaged over circles of diameter 0.8 degrees, has a distribution with width less than 0.7% (rms).
- "R-SIR-CAL-F-070 Instrument Response Stability (FoV)": Within each Wide Survey field, the relative response of each NISP detector should be known, such that fluctuations in the calibrated flux measurement between different detectors are smaller than 3% (rms) at any given time during the survey.

3.3 OU-SIM Simulations

In order to validate the whole *Euclid* pipeline before launch, detailed and complete simulations are needed. As described in Sec. 2.4 the Organisation Unit (OU) in charge of simulating *Euclid* data is OU-SIM. The goal of SIM is to develop raw Science Ground Segment simulations in a coordinated and coherent environment under the Science Data Centres framework for a massive production and distribution [56]. As it is shown in Fig. 3.4, before the launch, SIM provides all the necessary input raw data for the *Euclid* pipeline. These data are used by the OUs in the SGS to develop and validate their algorithms. The OU-SIM developers have written and are maintaining a software simulator for each *Euclid* observational channel, so there are mainly 3 different simulators. One is for VIS, another one for NISP in photometric mode (NISP-P) and the last one for NISP in spectroscopic mode (NISP-S). In this section we describe how the simulations are done by OU-SIM, with a particular focus on the spectroscopic channel.

3.3.1 The True Universe

All the instrument simulators in OU-SIM share a common input model of the Universe as it will be observed by *Euclid*, which is named *True Universe* (TU). Essentially it consists in a catalogue of astronomical sources, and it can be subdivided in two distinct catalogues: the star catalogue and the galaxy catalogue. The former contains the Milky Way stars that we expect to fall in the *Euclid* FoV. The latter collects the galaxies that are expected to be observed in

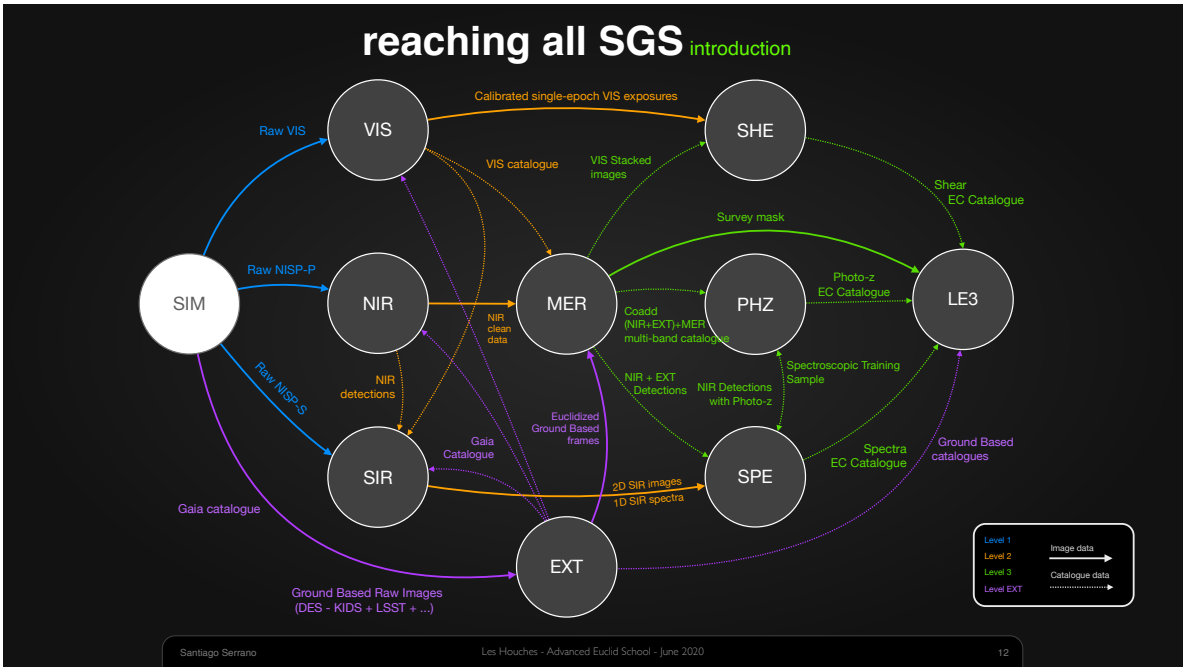

Figure 3.4: Scope of SIM in the SGS. Courtesy of Santiago Serrano [56].

the survey. Clearly both these catalogues are built by using models and simulations, since the currently available data are not sufficient to predict exactly what sources will be observed. In this section I briefly describe the True Universe star and galaxy catalogues that are used as an input for all the image simulators.

Star catalogue

Since *Euclid* will observe approximately one third of the whole sky, a wide stellar catalogue is needed in order to provide a realistic foreground for the simulated images, since the most interesting sources from the cosmological point of view are the galaxies. The stars that have to be simulated are located in our host galaxy, the Milky Way. At the moment not even the GAIA star catalogue [57] has enough stars for building a realistic stellar sky for *Euclid* simulations. In fact the maximum AB magnitude that is expected to be reached by *Euclid* is 24, while GAIA reached a maximum magnitude of 20.

A synthetic model is therefore needed to simulate the faint stars that are expected to be observed. The model that has been used is the Besançon model, which has been first developed in [58]. Most free parameters of this model are first constrained using solar neighbourhood data, and then extrapolations are made in order to predict the stellar population in more distant zones of the galaxy. For the bright stars it is important to accurately simulate their position, and for this reason the real data from the PICKLES catalogue [59] have been used. Also the Drimmel 3D model [60] is employed to take into account interstellar extinction.

For each star in the catalogue there is a specific set of columns containing all the needed pieces of information about the source. These columns can be essentially split into two main categories: *parameters* and *fluxes*. The parameters include: the source id, which is a unique

identifier for the source, the Right Ascension (RA) and Declination (DEC) coordinates, the radial distance to the star and the apparent magnitude in the Vega system. The fluxes category includes all the magnitudes of the source as observed from VIS and from NISP in the 3 photometric bans Y, J, H, together with the apparent magnitudes as are expected to be seen from the ground-based (EXT) observatories.

The spectra of the stars are also needed for the spectroscopic image simulations, together with the parameters listed above. A modified version of the Basel Spectral Energy Density (SED) library [61] has been employed with this purpose. The library provides stellar spectra templates, which are used to generate dispersed spectra as observed from NISP in the spectroscopic channel.

The TU star catalogue is made available to the SGS through a specific Data Product [1] named `DpdStarsCatalogProduct`: this is an `XML` file and contains the references to all the `FITS` files which make up the whole catalogue of sources, together with the coordinates of the pointings where the sources are expected to be observed. The catalogue files are inserted into the the Euclid Archive System (EAS), through which are accessible to the whole SGS.

Galaxy catalogue

The TU galaxy catalogue is built using the outputs of the *Euclid* Flagship simulation executed by the Cosmological Simulations Science Working Group, which is the official cosmological N-body simulation for *Euclid* [62]. This catalogue is essentially a table, with a specific set of columns for each galaxy. Additional pieces of information are needed for including the galaxies in the simulated images.

A morphology distribution based on [63] has been employed, in order to generate the images (thumbnails) of the galaxies to be simulated. All OU-SIM simulators share a common stamp library for generating galaxy thumbnails, which is based on the `GalSim` software [64], and includes both intrinsic galaxy shapes and weak lensing effects. Currently the light profile of the galaxies is based on a simple 2-Sersìc profile, but improvements and refinements to this model are foreseen. A SED template library based on COSMOS data [65] is used to generate the galaxy spectra to simulate the spectroscopic images, but some modifications have been made.

The intrinsic extinction laws coming from the flagship simulation have been applied, and a fixed set of emission lines has been added, with the line widths following the Tully-Fisher relation [66]. Among the emission lines there is obviously the Hα line, since it is be the main spectral feature used for measuring the spectroscopic redshifts. The calibration of the abundance and distribution of Hα-emitting galaxies is based on [67]. Finally the redshift of the source is applied to the galaxy spectrum, with both the cosmological and the peculiar velocity contributions. As for the star catalogue, the columns of the galaxy catalogue can be split into parameters and fluxes. The parameters include the source id, RA and DEC coordinates, together with the apparent coordinates shifted by gravitational lensing. Also the fluxes of the emission lines are present. Regarding the fluxes there is the same set of magnitudes that are present in the star catalogue.

The galaxy catalogue is distributed through a `DpdGalaxyCatalogProduct XML` file, and analogously to the star catalogue it contains the references to the `FITS` files which makes up

[1] A data product is an official file format for internal communication between the OUs, which is maintained and versioned by the OUs themselves and the SGS System Team.

the full table of galactic sources to be included in the simulations, which are accessible to the SGS through the EAS.

3.3.2 Mission Database and survey definition

Together with the source catalogues, the simulations need also all the information regarding the mission and the survey. The Mission Database (MDB) contains all the parameters of the mission: the optical models for distortions and the instrument models are hosted in the MDB, along with the dithering strategy and the maps of the background light. The MDB is a single `XML` file, which almost always refers to the name of the files containing the specific pieces of information. The MDB is a Data Product, and as such is versioned and updated multiple times during each Scientific Challenge. For this reason, a dedicated team is maintaining it.

The survey definition is contained into an `XML` file which hosts the list of the pointings of the telescope; for each pointing, the coordinates in the ecliptic system are reported, along with the attitude of the telescope. The survey file is available through the EAS and it is versioned and updated at each Scientific Performance Verification.

3.3.3 TIPS Simulator

TIPS (TIPS Is a Pixel Simulator) is the official OU-SIM code for simulating the NISP spectroscopic images to be used in the SGS. The software is mainly written in Python and simulates the images for each detector of the focal plane for a given observation sequence. The inputs for TIPS are essentially two:

- the True Universe catalogue for the specific pointing to be simulated;
- a corresponding catalogue of spectra for the sources to be simulated;
- the instrument configuration, specifying which instrumental and optical effects to include in the simulation, together with the paths to instrument model data.

The TIPS pre-processing prepares the sky model input to provide a list of images to be simulated. Each image simulation is then independent and could be run separately. The model of sky sources is defined with a catalogue of objects and a catalogue of spectra. The catalogue of object is derived from the True Universe catalogues for a given pointing and contains, for each source: the position on the sky, the shape parameters and a reference to a model spectrum in the catalogue of spectra. The model spectra can be given as input or generated on the fly based on the SED templates described in Sec. 3.3.1.

Image simulation

The simulation of the images is done with a modified version of aXeSIM, an image simulator developed initially for the Hubble Space Telescope [68]. For each source, aXeSIM computes the contribution on the image through each dispersion order of the grism. AXeSIM allows defining the PSF with an image, which is the best way of simulating for the performance analysis when the instrument is finalised. However during the conception phase, exploring the performance using a PSF image model is difficult, and a parametric model is often required. The simplest way to model different energy distributions in the PSF is to use a double Gaussian profile,

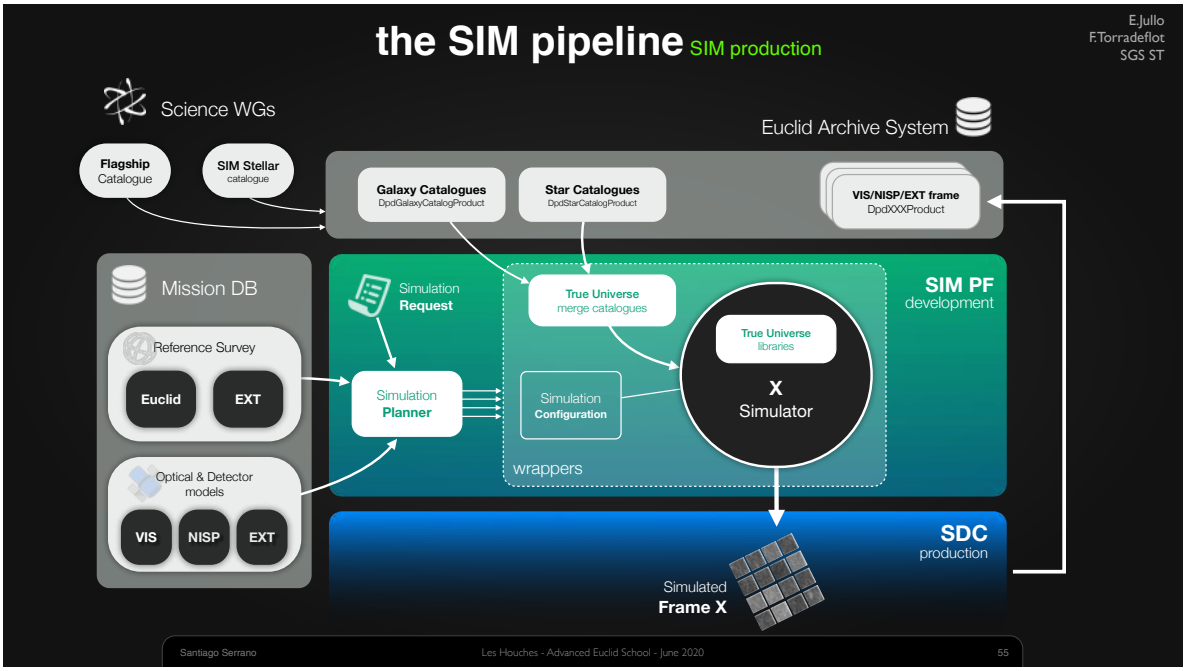

Figure 3.5: Schematic representation of the orchestration system designed by OU-SIM. Courtesy of [56].

which can describe a wide variety of energy distribution with only three parameters. AXeSIM also takes as input a distortion model to implement spectroscopic and astrometric distortions for the spectra. The result of this first step of the simulation is an image without detector noise, which is added a posteriori.

Detector simulation

The detector noises are added to the image computed above. The noises could be defined either with a single number or with a map. With a single number, all pixels have the same properties; with a map, different pixels may have different properties. The noise map model is clearly more realistic. Maps of quantum efficiency, dark current, and readout noise can be derived directly from the detector characterisation. In particular, the detector noise maps for TIPS are contained in the MDB, which refers to the `FITS` file storing the maps for each of the 16 NISP detectors. The usage of maps also allows us to include more complicated noise, like cosmic rays maps computed by another software.

The resulting output of each detector includes: a raw image, an error image, and a mask. The mask is being used to identify the dead pixels and those hit by cosmic rays.

3.3.4 The SIM Pipeline

As detailed at the beginning of this section, OU-SIM has to provide raw Science Ground Segment simulations for a massive production and distribution. For this reason a software pipeline has been designed as schematically represented in Fig. 3.5. The main entry point of

the pipeline is the so called Simulation Planner (`SimPlanner`), which is unique for all the three simulators. The `SimPlanner` receives the following inputs:

- a simulation request, containing specific information about the configuration for the simulation, comprising also the telescope pointings;
- the star and galaxy catalogue data products described in Sec. 3.3.1;
- the Mission Database (MDB);
- the survey definition file.

For each pointing the `SimPlanner` produces the specific configuration file for the particular simulator, denoted by a generic X in Fig. 3.5, together with the TU for the particular sky region observed in the simulation. This system also implements the dithering strategy, which is described in the MDB, and accounts for uncertainty on the positioning of the telescope. The production of simulations is done within a Science Data Centre (SDC), where it can be parallelised.

After an official SIM pipeline run has finished, only the simulated raw images are ingested into the EAS, while all intermediate products are not stored. The procedure just described presents several advantages when dealing with massive production and distribution. Nonetheless, the procedure lacks an high level of control and flexibility which could be needed for some particular validation procedures of the scientific software pipelines, like the SIR pipeline described in Sec. 3.2, which uses simulated data as input. In the next section I report the results of my work on the production of the high-control simulations needed for the SIR Pipeline validation.

3.4 The `SIR_SpectroSim_Runner` software

The OU-SIM software pipeline orchestrates automatically the simulations production, making the simulated images as realistic as possible, as reported in the last section. After having chosen the telescope pointings to simulate, the software automatically selects all the sources which are expected to be observed in the FoV from the star and galaxy catalogues, and includes all the available detector noise effects from the MDB.

The simulations provided by OU-SIM are as realistic as possible, which implies that all the systematic and cross-contamination effects are included. The input spectra used for generating the simulated images are not stored as output files, since these are not official data products of the SIM pipeline. It is therefore impractical or impossible to test the scientific SIR pipeline described in Sec. 3.2 using the OU-SIM pipeline. However, spectroscopic simulations are needed for the science performance verification, to test if OU-SIR fulfills its scientific requirements and, from a more practical point of view, to spot bugs in the code. A step by step validation, where all the pipeline blocks are tested singularly in an isolated way, is only possible with full control on the simulation side. It is in particular necessary to have the freedom to choose

- the source catalogue to be simulated in the FoV, as well as their theoretical spectra;
- the instrumental and optical effects included in the simulation.

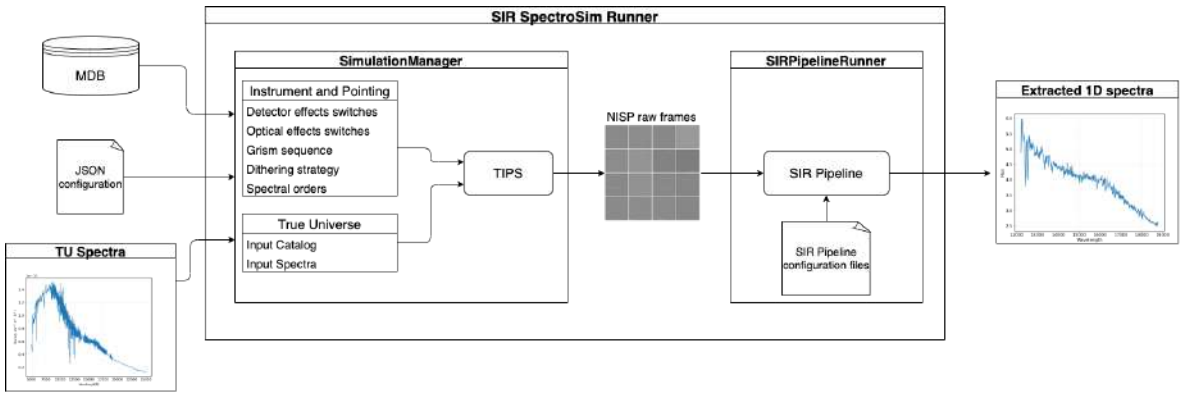

Figure 3.6: Block diagram of the SIR_SpectroSim_Runner software. Starting from a single JSON file and all the necessary data pointed by MDB, it orchestrates the run of TIPS and/or the SIR Pipeline.

In the first part of my PhD I worked on the development of a software element aimed at performing the customised and granular spectroscopic simulations just described. My code stands on top of TIPS, the official spectroscopic simulator described in Sec. 3.3.3, and it allows us to control it in a standalone mode, disentangling it from the SIM pipeline outlined in Sec. 3.3.4. The code I wrote has been given the name SIR_SpectroSim_Runner, and it is available [2] through the official gitlab of the SGS

https://gitlab.euclid-sgs.uk/lpagan01/SIR_SpectroSim_Runner

The whole software package has been written in Python, and consists in a total of 2775 lines of code, excluding blank lines and comments. The code has been used in the OU-SIR group for automating the validation procedures, allowing also to control the results from end to end. A great effort has been made to keep the simulation production as simple as possible, minimising the amount of configuration and input files needed by the software (see Fig. 3.6). SIR_SpectroSim_Runner allows us also to run the SIR Pipeline data reduction on the simulated images, and it is therefore made by two distinct modules: the SimulationManager and the SIRPipelineRunner. As the name tells, the simulation side of the code is managed by the SimulationManager module. The configuration for the simulation is specified through a JSON file, which contains:

- boolean switches for detector and optical instrumental effects;
- the grism observation sequence;
- the dithering strategy, with custom sky coordinates offsets;
- the name of the file containing the True Universe (TU) input source catalogue;
- the name of the file containing the True Universe input spectra for the sources specified in the TU catalogue.

[2] At the moment the access to the SGS gitlab code is limited to Euclid Consortium members only.

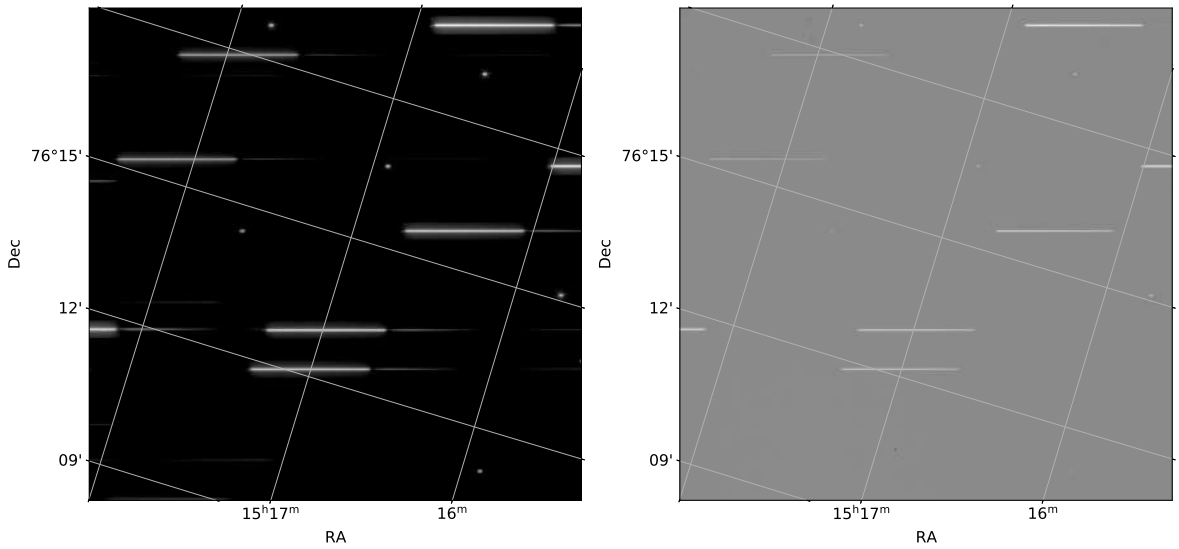

Figure 3.7: Comparison between a simulated NISP detector without (left) and with (right) background light. This simulation has been made using the `SIR_SpectroSim_Runner` software, and only bright stars (magnitude in NISP J band < 15) from True Universe are present in the FoV.

The `SimulationManager` allows us to simulate one, two, three, or four dither exposures for the instrument. Additional input files are represented in Fig. 3.6 by the MDB data, which must be provided in a working directory whose path can be specified in the `JSON` configuration. After reading the configuration, the `SimulationManager` creates all the necessary input files for TIPS and then it starts the simulation. After the simulation has ended, the `SimulationManager` reorganises the working directory, optionally storing all the necessary intermediate products which can be used for validation. The intermediate files include the input spectra falling on each given detector and the instrument configuration file used by TIPS to simulate the images.

The `SIRPipelineRunner` module manages the run of the SIR Pipeline on the simulated raw frames produced by the `SimulationManager`. Running the SIR Pipeline is not a straightforward task. As represented in Fig. 3.3, the SIR Pipeline is made by several logical blocks, each of them taking as input the output of the previous one. The `SIRPipelineRunner` orchestrates the pipeline run, making sure that there is consistency in the whole process. For example if the background light is not present in the simulated images, the runner skips the background subtraction task. Moreover the `SIRPipelineRunner` provides a configuration file to each task of the pipeline, as Fig. 3.6 illustrates.

In the next section I report some of the several test cases in which `SIR_SpectroSim_Runner` has been used to validate the SIR Pipeline.

3.5 SIR Validation test cases

The `SIR_SpectroSim_Runner` code has been used as the main tool for the validation of the SIR pipeline at all levels. The granularity of the simulations produced allows to test what is the impact of every single instrumental systematic effect. Also, the full control on the

Figure 3.8: Flux-magnitude relation for extracted spectra in a simulation with only bright stars and no background. The flux on the y axis is measured in photon counts. Two different linear trends are clearly visible in the counts-magnitude relation. Plot courtesy of Marco Scodeggio.

catalogue of sources simulated in the FoV gives the possibility to avoid cross-contamination among spectra, which is one of the most difficult noise sources to handle. Moreover, the possibility to automatically run the full data reduction after the simulation makes it possible to perform end-to-end tests, where the extracted spectra are directly compared to the input theoretical ones. In this section I report some of the test cases of the SIR Pipeline for which the `SIR_SpectroSim_Runner` code has been employed.

3.5.1 Finding bugs in the code: spectra resampling

The first test case I present is one where `SIR_SpectroSim_Runner` has been helpful in finding a bug in the SIR Pipeline. The simulation employed in this case was one with only a few bright stars and without background light (see left panel of Fig. 3.7). A problem was spotted by plotting the extracted spectra counts against the magnitude of the sources. If things are simulated properly and the spectra are correctly extracted, the logarithm of the flux of the sources should be linearly proportional to their magnitude. However, as Fig. 3.8 shows, this was not the case. In the plot, each red dot represents a source in the FoV whose spectrum had been extracted by the SIR Pipeline. There are clearly two different linear trends in the flux-magnitude relation of the extracted spectra, instead of one as it should be. The flux of some sources was systematically being underestimated[3].

Going in more depth with additional tests, it turned out that the SIR Pipeline was somehow treating differently the sources, depending on the fractional part of their vertical coordinate. The problem is manifest in the plots shown in Fig. 3.9. The x axis is the fractional part of the

[3]In principle there could also have been a spurious overestimation, but this turned out not to be the case.

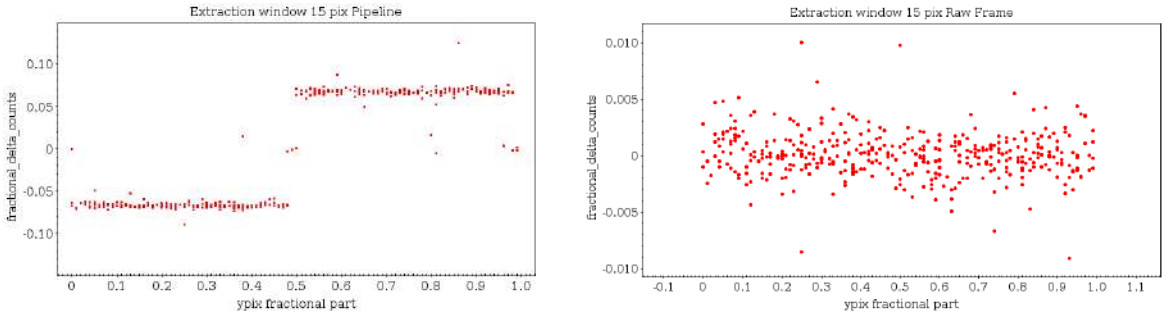

Figure 3.9: Fractional difference on the fluxes of the sources in two different exposures against the fractional part of the vertical coordinate. Left panel refers to the fluxes as extracted from the SIR Pipeline. Right panel refers instead to the fluxes computed by summing up the values of the pixels of the two-dimensional spectra *before the extraction*. Plots courtesy of Marco Scodeggio [69].

vertical coordinate of the sources. The y axis represents the fractional difference of the fluxes of the sources between two exposures, in which the same sources have been translated by 0.5 pixels along the vertical (orthogonal to dispersion) direction. The difference is normalised to the mean of the fluxes in the two situations, in order to be unbiased about which flux was the correct one. On the left panel, the fluxes involved are those extracted by the Pipeline, while on the right panel the fluxes were obtained by summing up the value of the pixels of the 2-D spectra before resampling them. In both cases the extraction aperture for computing the fluxes was very broad (15 pixels), and it was possible to use such a large aperture since the simulated spectra were a few and isolated from each other, and no background was present in the simulation.

The total fluxes of the extracted spectra exhibited a non-physical dependence on the fractional part of the vertical coordinate of the sources. In particular, when the centre of the spectrum of the source fell in the upper half of a pixel, the extracted flux was systematically higher by about 10% with respect to the situation where it fell in the lower half. From the right panel it is clear that this feature was not present when computing the fluxes by summing up the pixels of the 2-D images. This implies that the simulated data were consistent and that the problem could only be in the extraction procedure. Since this behavior is not physical, some artificial effect induced by a bug in the software was likely to be present. Moreover this difference on the extracted flux was well above the required accuracy on the flux estimation reported in Sec. 3.2, and therefore the bug had to be found and corrected.

In order to understand what was happening it is useful to outline the details of the 2-D resampling procedure:

1. The resampling employs a 2-D kernel which covers a square window of 4×4 pixels.

2. Since the images of the 2-D spectra are cropped, the resampling kernel may enclose some pixels that are outside of the cropping window, and the code automatically assigned 0 as the fill value for these missing pixels. Instead it should have ignored them, and here there was the error.

3. The largest contribution from the resampling comes from the centre of the spectrum,

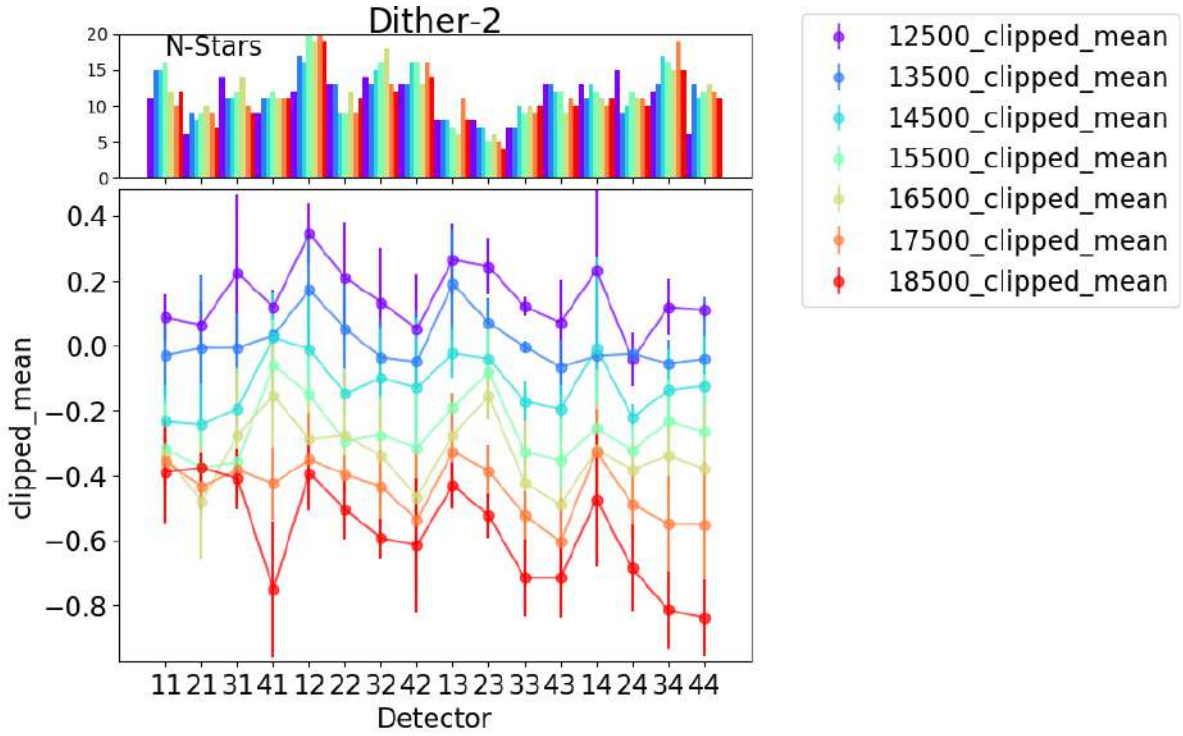

Figure 3.10: Difference (in pixel) between the input and reconstructed peak of the spectrum, averaged over the sources falling in each NISP detector. The rainbow colormap represents the gradual increasing of the wavelength, from the bluest (12 500 Å) to the reddest (18 500 Å) part of the spectrum. Plot courtesy of Chiara Mancini and Marco Scodeggio [69].

since it is the brightest part. Then in order to correctly estimate the extracted flux, it is necessary to correctly resample the image especially around the centre of the spectrum.

4. When the centre of the spectrum happened to be in the upper half of the pixel, the resampling kernel did not go outside of the extraction aperture, therefore not including spurious zero values. When the centre happened to be in the lower half instead, the kernel window included some spurious zeros, effectively lowering the resulting flux.

This error could only have been spotted with a very special simulation with few bright sources and no background. In such a simplified simulation a very wide extraction aperture can be used without contaminating the extracted spectra with other sources or with background light. Instead, in a complete simulation this non-physical effect would have been buried by the high cross-contamination between the thousands of sources and all the background effects, and therefore it could not have been spotted.

3.5.2 Tuning pipeline performance: verifying grism tilt reconstruction

When the grism is tilted the dispersion direction is neither horizontal nor vertical, and this must be considered in the data reduction. The extraction window is rotated by an angle which has to be as close as possible to the true unknown grism tilt, which has to be estimated from

the images. In the SIR Pipeline there is a dedicated algorithm for the tilt estimation, based on a two-dimensional Fourier transform, which must be validated.

The validation of the grism tilt reconstruction starts by simulating clean images with few well separated objects, not subject to cross-contamination, with `SIR_SpectroSim_Runner`. The validation procedure tests the accuracy of the reconstruction of the trace left by the spectrum, which is represented by the curve followed by the brightest point of the spectrum as a function of the horizontal pixel coordinate. With a clean simulation it is possible to measure the brightest point of the spectrum as the maximum of the 2D spectrum along the vertical coordinate at various wavelengths. The resulting trace is then compared to the trace reconstructed by the SIR Pipeline, in which also the astrometric and spectroscopic distortions are taken into account. Therefore this test also validate the implementation of the curvature model implemented in the pipeline, which parameterises the spectroscopic distortions.

In Fig. 3.10 the results of the validation are reported. The plot shows the difference between the input and the reconstructed peak position, averaged over all the sources falling in each of the 16 detectors. This has been done for seven different wavelengths between 12 500 Å and 18 500 Å, in order to probe the accuracy of the reconstruction locally along the dispersion direction. From the plot it can be seen that there is a slight systematic difference, which gradually increases when going from the blue to the red part of the spectrum. This may indicate that there has been a systematic effect in the tilt reconstruction, or it may simply represent the resolution in the trace reconstruction of the algorithm. Other software tests including various telescope pointings have been scheduled, in order to distinguish between the two possible explanations.

3.5.3 Optimising pipeline parameters: extraction aperture

Simulations can also be used to determine the optimal values of some parameters of the data reduction. One of these parameters is the extraction aperture, which should be as large as it is necessary to enclose the whole spectrum, and thus maximise the desired signal. However, a wider aperture increases the contribution from the background, which is approximately constant in the FoV. This therefore decreases the signal-to-noise ratio, which in this context is defined as

$$\mathrm{S/N} = \frac{\text{enclosed spectrum flux}}{\text{rms of background flux}} \tag{3.3}$$

The enclosed spectrum flux converges at the value of the total flux of the object when the extraction window encloses the whole spectrum. The enclosed background flux instead grows approximately linearly with the extraction aperture, since the background flux is approximately constant in the FoV. The S/N ratio starts at zero, when the aperture is zero. This is the degenerate case where no flux is enclosed in the extraction window. Then the S/N rises reaching a maximum, corresponding to the optimal aperture for the given object. Finally the S/N decreases after the enclosed flux object converges to the total flux of the source. In this regime, the background contribution dominates, since the signal flux remains constant, and the enclosed background flux grows.

The optimal aperture is found with an empirical procedure, using multiple reductions of the same simulation with different apertures. However this procedure can only converge with meaningful results, if the simulated spectra are isolated and not contaminated from overlapping

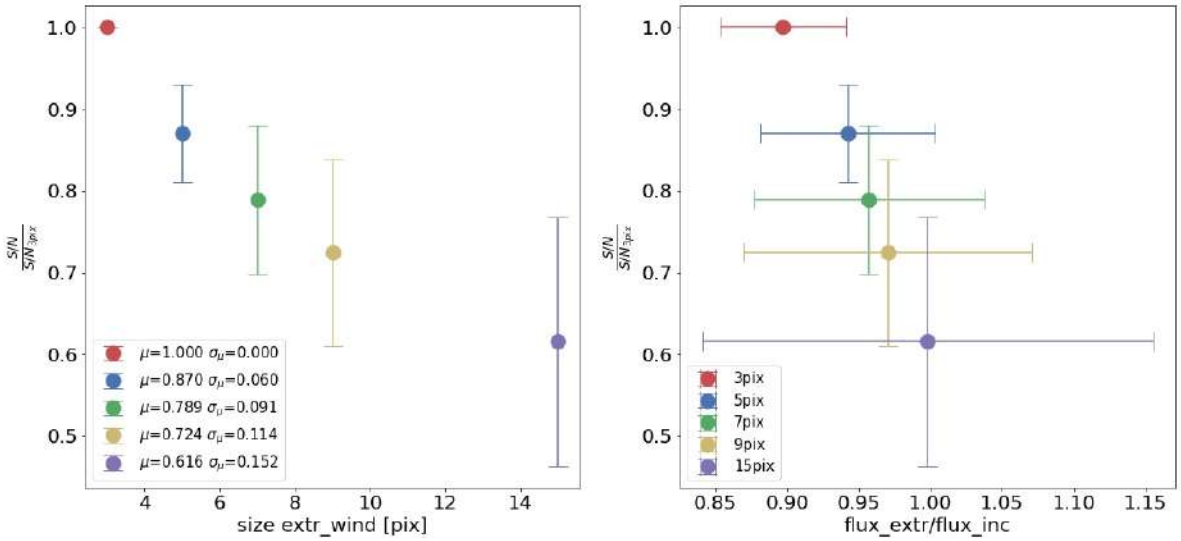

Figure 3.11: Signal-to-noise S/N ratio against the extraction window (left) or the extraction efficiency (right), defined as the ratio of the extracted flux over the incident flux. The plots show that increasing the aperture lowers the S/N, but it also increases the extraction efficiency. Plots courtesy of Chiara Mancini and Marco Scodeggio [69].

sources. This means that a completely realistic simulation cannot be used for the purpose. Instead, a simulation with a few bright objects [4] with the background enabled are needed.

An image of one detector given from the simulation in exam is shown in the right panel of Fig. 3.7. In Fig. 3.11 the results of the optimisation procedure are shown. In the left panel the S/N relative to its value at 3 pixel aperture is plotted against the aperture size. The points represent averages over the population sources in the field of view, with error bars given by the standard deviations. As expected, the S/N decreases with the aperture size, and the maximum is at 3 pixel aperture. However, also the fraction of enclosed spectrum flux must be considered to determine the optimal aperture. In the right panel the same S/N has been plotted against the ratio between the extracted and the input spectrum flux, referred to as the *extraction efficiency*. Using an aperture of 3 pixels leads to the best S/N, at the cost of losing 10% of the spectrum flux. The loss is also larger in the case of extended objects like galaxies, which are the most interesting sources for *Euclid* science. Instead, choosing an aperture between 5 and 7 pixels yields a S/N which is $\sim 80\%$ of the maximum, along with an extraction efficiency greater than 95%. For this reason it has been concluded that an aperture between 5 and 7 pixel is optimal, since it leads to reasonably high S/N and extraction efficiency.

3.5.4 Performing end-to-end tests: the full simulation-reduction chain

The architecture of the `SIR_SpectroSim_Runner` software represented in Fig. 3.6 shows how it is possible to perform end-to-end tests, in which the whole software pipeline is tested. The user is free to choose the input spectra, which can be used as a benchmark to check the quality of the output extracted spectra. Moreover the high-level of automation in the whole procedure

[4] In this simulation only stars with J-band magnitude magJ in the range $14 < \text{magJ} < 18$ have been included.

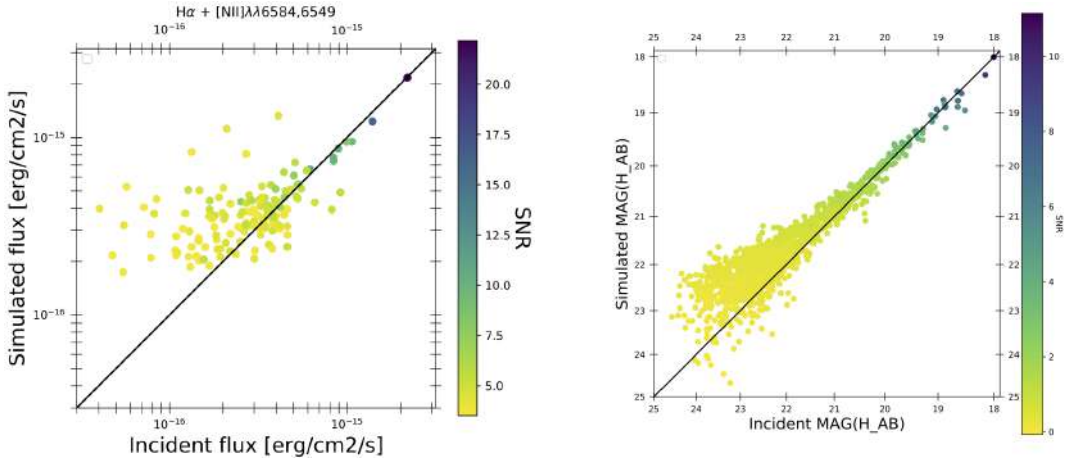

Figure 3.12: Plots of the extracted fluxes against the incident (theoretical) ones for the check described in Sec. 3.5.4. In the left panel the extracted Hα line fluxes are compared to their theoretical counterparts, while the right panel shows the same plot but for the continuum of the spectra. Plots courtesy of Louis Gabarra [70].

makes it easy to perform such a task.

An end-to-end test that has been performed is a consistency check on the Hα line flux for some simulated galaxies, whose spectra were taken as input. This line flux is defined as the integral of the galaxy spectrum over a suitably defined range centred at the Hα line *observed* wavelength. Correctly reconstructing the Hα flux is of paramount importance, since this is the main spectral feature that the SPE pipeline uses for measuring the spectroscopic redshifts. Significant discrepancies in the estimated line flux can affect the completeness and the purity of the catalogue, which are of major importance for the scientific requirements, as explained in Sec. 2.3.3.

In order to be in the spectroscopic catalogue, the galaxies must have an Hα line falling within the sensitivity region of NIR detectors, which is $1100 - 2000\,\mathrm{nm}$ as reported in Chapter 2, and with an intensity above the sensitivity threshold of the *Euclid* deep survey. The first condition is effectively a constraint on the redshift of the source, which must be comprised in the range $0.9 < z < 1.8$. The second condition translates into an Hα line flux $> 5 \times 10^{-17}\,\mathrm{erg\,cm^{-2}s^{-1}}$. The end-to-end test has been performed with a simulated ad hoc galaxy catalogue, using as sources only the galaxies satisfying these conditions, and positioned on a regular grid. The simulation excluded all the irrelevant sources, useless for the test itself, thus gaining a significant improvement in the pipeline processing time. The spacing between the galaxies has been chosen large enough to prevent cross-contamination, since this effect was not relevant for this particular check.

In Fig. 3.12 two plots of the extracted flux against the incident (theoretical) one are shown. In the left panel the Hα line flux from extracted spectra are plotted against their simulated value, computed from the input galactic spectra. In the right panel instead, the plot has been repeated for the continuum[5] of the galaxy spectra. In both cases the black diagonal line represents the identity relation, which is the expected ideal result for this test. In both

[5]Here with continuum it is meant the broad, slowly-varying component of the spectrum.

cases the scattered points cluster around the black diagonal, which means that there is a good agreement between the simulated input spectra and the extracted output ones. We also notice an asymmetric behaviour with respect to the diagonal at low values of the incident fluxes. The asymmetry is particularly evident in the case of Hα line flux, reported in the left panel of Fig. 3.12, and it starts when the incident flux is around the line flux limit of the *Euclid* wide survey, i.e. $2 \times 10^{-16}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$.

The asymmetry has been attributed to the influence of the background light, which was present in the simulation. In a spectroscopic exposure of the *Euclid* wide survey the detector counts coming from background infrared photons are expected to follow a Poisson statistic with a mean value of ~ 900, which is approximately uniform across the field of view. Consequently the standard deviation of pixel-to-pixel stochastic fluctuations of the background counts can be estimated to be about $\sqrt{900} = 30$. As explained in Sec. 3.2, in the SIR Pipeline there is a step in which the background is subtracted from the input frames. However this subtraction only removes the average value of the background as inferred from the raw frames, while its stochastic pixel-to-pixel fluctuations cannot be eliminated.

On the other hand the typical Hα counts of a galaxy near the flux limit are ~ 100, with a standard deviation of $\sqrt{100} = 10$, therefore for these sources the background fluctuations represent a dominant contribution. Moreover the Hα line photons counts are distributed in a localised region on the detector of approximately 3×3 pixels. On such a small region, the random pixel-to-pixel background fluctuations may lead to a significant random variation of the observed line flux. These fluctuations of the background may either increase the counts of a faint source above the detection threshold, or decrease them below the threshold. In the first case, the extracted flux is systematically higher than its simulated incident flux. In the second case, the source is not detected. This explains the high asymmetric scatter at low flux observed in the distribution of the galaxies represented in Fig. 3.12.

In the case of the continuum of the spectra reported in the right panel of Fig. 3.12 the low-flux asymmetry due to the background fluctuations is less prominent. This arises because the continuum flux is estimated by using a larger portion of the spectra, and the average contribution of background fluctuations becomes close to zero.

Chapter 4

Exploiting multi-probe correlations in *Euclid* forecasts

The *Euclid* mission will be one of the largest galaxy surveys ever made, covering about one-third of the sky. As explained in Chapter 2, *Euclid* will produce visible imaging data of 1.5 billion galaxies for weak lensing measurements. At the same time the redshifts of these galaxies will be estimated by using NISP in photometric mode. Instead, with NISP in spectroscopic mode the spectroscopic redshifts of about 20 million galaxies will be measured with a much higher precision, using the Hα emission line feature of their spectrum. Therefore, *Euclid* represents a unique opportunity to exploit a full combined analysis of three cosmological probes: weak lensing (WL), photometric galaxy clustering ($\mathrm{GC_{ph}}$), and spectroscopic galaxy clustering ($\mathrm{GC_{sp}}$). On the one hand, the statistical *cross-correlations* between these probes potentially contain new cosmological information, which may enhance the final constraints on the cosmological parameters. On the other hand, these cosmological probes can also exhibit a non-zero cross-covariance, which if not accounted for may lead to an overestimation of the constraining power coming from their combination.

A full data analysis including all the cross-correlations and cross-covariances between the probes may be computationally expensive. Therefore *forecasts* are needed in order to get insights on what would be their impact on the final scientific results. In [5] a first Fisher forecast analysis – which in the following is also referred to as IST:F – for the *Euclid* survey has been done. In that work, the angular spherical-harmonic measurements of the two-point statistic (which from now on will be called harmonic power spectrum for simplicity) was adopted as the observable for WL and $\mathrm{GC_{ph}}$, while the the full, anisotropic, and redshift-dependent galaxy power spectrum (which from now on will be called Fourier power spectrum for simplicity) was used as the observable for $\mathrm{GC_{sp}}$. The authors studied the impact of the $\mathrm{XC(WL, GC_{ph})}$ cross-correlation, showing that its inclusion enormously improves the constraints on cosmological parameters. In particular, it increases the dark energy FoM by about 400% in the optimistic scenario, a result which has been reproduced in the present work too. The cross-correlations and cross-covariances involving $\mathrm{GC_{sp}}$ and the other probes were not considered, since it was not obvious how to evaluate the correlation between a probe studied in the three-dimensional (3D) Fourier domain with another treated in the two-dimensional (2D) harmonic one.

The aim of the present work is to extend the IST:F forecast analysis with the inclusion of the WL-$\mathrm{GC_{sp}}$ and $\mathrm{GC_{ph}}$-$\mathrm{GC_{sp}}$ covariances and cross-correlations, quantifying their impact on

the parameter constraints. In order to do this, the 2D harmonic formalism is adopted for all the probes (harmonic approach) including $\mathrm{GC_{sp}}$. In this way the cross power spectra can be easily computed and included in the analysis. Also, it is possible to perform the so-called full 6×2pt analysis, which includes as observables all the six two-point correlations functions that can be constructed from the three cosmological probes: WL, $\mathrm{GC_{ph}}$, and $\mathrm{GC_{sp}}$[1]. Furthermore, with this approach all the possible covariances can be naturally included in the analysis. Nonetheless, this approach presents a drawback, consisting in the fact that $\mathrm{GC_{sp}}$ significantly loses much of its constraining power when projected in the 2D harmonic domain from the 3D configuration space. In order to circumvent this issue, two different strategies have been adopted. The first is to use as observable for the $\mathrm{GC_{sp}}$ auto-correlation the harmonic power spectrum, but refining the tomographic binning in order to enhance its performances. The second is to consider an alternative *hybrid* approach, which tries to combine the information coming from the harmonic $\mathrm{GC_{sp}}$ cross-correlations with the full constraining power of $\mathrm{GC_{sp}}$ auto-correlation expressed as the 3D galaxy power spectrum. This is done by computing the cross-correlation power spectra $\mathrm{XC(WL, GC_{sp})}$ and $\mathrm{XC(GC_{ph}, GC_{sp})}$ in the *harmonic* domain, and then adding the Fourier auto-power spectrum of $\mathrm{GC_{sp}}$ – which in the following is referred to as $\mathrm{GC_{sp}}(P_k)$ – as an independent probe. However, this approach is approximate: it implicitly neglects the cross-covariances between the auto-spectrum $\mathrm{GC_{sp}}(P_k)$ and the harmonic spectra of the other probes, since the Fisher matrix associated to $\mathrm{GC_{sp}}(P_k)$ is added *a posteriori* in the covariance as an independent probe.

The forecast has been performed in two scenarios: one pessimistic and one optimistic, differing only for the multipole ranges used to compute the harmonic power spectra. For a summary of the settings see Tab. 4.5. The results of the analysis show that the harmonic-domain cross-covariances between $\mathrm{GC_{sp}}$ and the other two probes, WL and $\mathrm{GC_{ph}}$, are always practically negligible. The $\mathrm{XC(GC_{ph}, GC_{sp})}$ and $\mathrm{XC(WL, GC_{sp})}$ cross-correlations start to give a slightly significant impact only when using a very fine tomographic binning for $\mathrm{GC_{sp}}$, but their contribution never reaches the impact given by the $\mathrm{XC(WL, GC_{ph})}$ cross-correlation. It turned out that the constraining power of $\mathrm{GC_{sp}}$ treated in the harmonic domain is dramatically limited by the higher shot noise with respect to $\mathrm{GC_{ph}}$. In fact the *Euclid* spectroscopic sample is expected to contain a factor of ~ 80 less galaxies than the photometric one. Consequently, the $\mathrm{GC_{sp}}$ harmonic power spectra will be affected by higher shot noise than the $\mathrm{GC_{ph}}$ ones. Furthermore $\mathrm{GC_{sp}}$ covers a smaller redshift range than $\mathrm{GC_{ph}}$, and this results in a further reduction of its constraining power in harmonic domain with respect to $\mathrm{GC_{ph}}$.

The computations presented in this chapter have been performed using a code named `SEYFERT`, whose I have been one of the main developers. In Appendix B I describe the overall architecture of the code, together with the numerical algorithms implemented for carrying out the calculations. The results obtained with `SEYFERT` for WL, $\mathrm{GC_{ph}}$, and $\mathrm{XC(WL, GC_{ph})}$ have been validated against the ones obtained by the codes used in the IST:F [5].

This chapter is structured as follows.

In Sec. 4.1 the harmonic power spectra formalism is described, together with a brief description of the Fourier power spectrum method used in the IST:F for treating $\mathrm{GC_{sp}}$. The cosmological model assumed is the flat w_0w_aCDM, where the dark energy equation of state is governed by the CPL parameterisation Eq. (1.122).

[1] Using three different probes three auto-correlations and three cross-correlations can be constructed, yielding a total of six independent 2-point functions.

Parameter	$\Omega_{\rm b,0}$	$\Omega_{\rm m,0}$	w_0	w_a	h	$n_{\rm s}$	σ_8	$\sum m_\nu$[eV]
Reference value	0.05	0.32	-1	0	0.67	0.96	0.816	0.06

Table 4.1: Values of the cosmological parameters in the reference cosmology considered in the forecast. All the parameters are let free to vary except for the sum of the neutrino masses, which has been fixed to its reference value $\sum m_\nu = 0.06\,\text{eV}$.

In Sec. 4.2 the general Fisher matrix formalism is reported. The aforementioned harmonic and hybrid approaches for computing the Fisher matrices are described in detail. Also, some examples of two-probe combinations are presented in order to concretely explain the effects of cross-covariance and cross-correlation on the Fisher matrix.

In Sec. 4.3 the results of the analysis are presented. The impact on the results of the cross-covariances and cross-correlations involving $\text{GC}_{\rm sp}$ is discussed in detail. This is done both for the pairwise combinations $\text{GC}_{\rm ph}$-$\text{GC}_{\rm sp}$, WL-$\text{GC}_{\rm sp}$ and for the full combination of $\text{WL}, \text{GC}_{\rm ph}, \text{GC}_{\rm sp}$ considered altogether.

Finally, Sec. 4.4 contains the conclusions of this work, along with future perspectives.

4.1 Observables

As said in the introduction, the cosmological model adopted in this forecast is a flat w_0w_aCDM cosmology. Therefore, the redshift evolution of the Hubble parameter is governed by Eq. (1.124) with $\Omega_{k,0} = 0$, and consequently $\Omega_{\rm DE,0} = 1 - \Omega_{\rm m,0}$. The cosmological parameters involved in the forecast are summarised in Tab. 4.1, which also reports the values in the reference cosmology.

The probes considered in this forecast are: the photometric galaxy clustering ($\text{GC}_{\rm ph}$), the spectroscopic galaxy clustering ($\text{GC}_{\rm sp}$), and the weak lensing (WL). For the WL and $\text{GC}_{\rm ph}$ probes, the observable employed is the tomographic angular power spectrum $C_{ij}(\ell)$. For the $\text{GC}_{\rm sp}$ probe, both the harmonic power spectrum and the Fourier power spectrum have been considered as observables. The Fourier power spectrum approach is the same followed in [5] and briefly described in Sec. 4.1.3. However, in this work the corresponding Fisher matrix is not directly computed; instead it is taken as an external input from the IST:F public online repository [71].

The angular power spectrum $C(\ell)$ is essentially the harmonic transform of the two-point tomographic angular correlation function. Each cosmological probe A in a given redshift bin i can be associated to a field $f_i^A(\hat{\mathbf{n}})$ defined on the sky, which can be expanded in spherical harmonics

$$f_i^A(\hat{\mathbf{n}}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m}^{Ai} Y_{\ell m}(\hat{\mathbf{n}})\,. \tag{4.1}$$

The $a_{\ell m}^{Ai}$ coefficients are the projection of f_i^A on the spherical harmonic orthonormal basis

$$a_{\ell m}^{Ai} = \int \mathrm{d}\Omega\, Y_{\ell m}^*(\hat{\mathbf{n}}) f_i^A(\hat{\mathbf{n}}) \tag{4.2}$$

The tomographic angular correlation between probe A, at the redshift bin i, and B, at the

redshift bin j, is then defined as the following implicit relation

$$\left\langle a_{\ell m}^{Ai} \left(a_{\ell' m'}^{Bj}\right)^* \right\rangle = C_{ij}^{AB}(\ell)\delta_{\ell\ell'}\delta_{mm'}\,, \tag{4.3}$$

where the angular brackets here denote the theoretical expectation value. In the Limber approximation the expression of the angular power spectra can be parameterised as

$$C_{ij}^{AB}(\ell) \simeq c\int_{z_{\rm min}}^{z_{\rm max}} {\rm d}z \frac{W_i^A(z)W_j^B(z)}{H(z)\chi^2(z)} P_{\delta\delta}\left[k=\frac{\ell+1/2}{\chi(z)}, z\right]\,, \tag{4.4}$$

where $\chi(z)$ is the comoving distance defined in Eq. (1.36), W_i^A is the weight (or window) function for the probe A in the i-bin (details are given in the next subsections), and $P_{\delta\delta}$ is the total matter power spectrum introduced in Eq. (1.108). In this forecast, it has been taken as an external input, loading the publicly available power spectra computed in [5]. These power spectra have been computed using the `CAMB` Boltzmann code with a revised version of the HALOFIT [72] recipe to take into account non-linearities. Massive neutrinos have been included among the Boltzmann code configuration parameters, using $\sum m_\nu = 0.06\,{\rm eV}$ as reference value.

The power spectra are publicly available at [73].

The functional form of the weight function W_i^A depends on the probe A. The fundamental ingredient for the computation of W_i^A is the redshift distribution per unit solid angle, ${\rm d}N^A(z)/({\rm d}z{\rm d}\Omega)$, associated to the probe. This can be either modelled analytically, as in the case of WL or $\mathrm{GC_{ph}}$, or can be obtained via simulations, as for $\mathrm{GC_{sp}}$.

The normalised redshift density of the probe A in the i-bin can be computed from the redshift distribution as

$$n_i^A(z) = \frac{\int_{z_i^-}^{z_i^+} {\rm d}z_{\rm p}\, \frac{{\rm d}N^A}{{\rm d}z{\rm d}\Omega}(z) p_A\left(z_{\rm p}|z\right)}{\int_{z_{\rm min}}^{z_{\rm max}} {\rm d}z \int_{z_i^-}^{z_i^+} {\rm d}z_{\rm p}\, \frac{{\rm d}N^A}{{\rm d}z{\rm d}\Omega}(z) p_A\left(z_{\rm p}|z\right)}\,. \tag{4.5}$$

The function $p_A(z_{\rm p}|z)$ is the probability that a galaxy with true redshift z will be measured with a redshift $z_{\rm p}$. This probability incorporates the redshift measurement errors, and it is modelled differently depending whether the redshift measurement is photometric or spectroscopic. The parameterisation chosen is the same of [5]

$$p_A(z_{\rm p}|z) = \frac{1-f_{\rm out}}{\sqrt{2\pi}\sigma_{\rm b}(1+z)} \exp\left\{-\frac{1}{2}\left[\frac{z-c_{\rm b}z_{\rm p}-z_{\rm b}}{\sigma_{\rm b}(1+z)}\right]^2\right\} + \frac{f_{\rm out}}{\sqrt{2\pi}\sigma_{\rm o}(1+z)} \exp\left\{-\frac{1}{2}\left[\frac{z-c_{\rm o}z_{\rm p}-z_0}{\sigma_{\rm o}(1+z)}\right]^2\right\}\,. \tag{4.6}$$

This model allows us to include both a multiplicative and additive bias in the redshift determination of a fraction $1-f_{\rm out}$ of sources with well measured redshifts, and a fraction $f_{\rm out}$ of catastrophic outliers, i.e. systems with severely incorrect estimate of the redshift. The values of the parameters for the probability functions are summarised in Tab. 4.2. The uncertainty $\sigma_{\rm b}$ on the correctly measured (not catastrophic) redshifts has been chosen from the *Euclid* design requirements [46] described in Chapter 2. In particular $\sigma_{\rm b} = 0.05$ for $\mathrm{GC_{ph}}$ and $\sigma_{\rm b} = 0.001$ for $\mathrm{GC_{sp}}$.

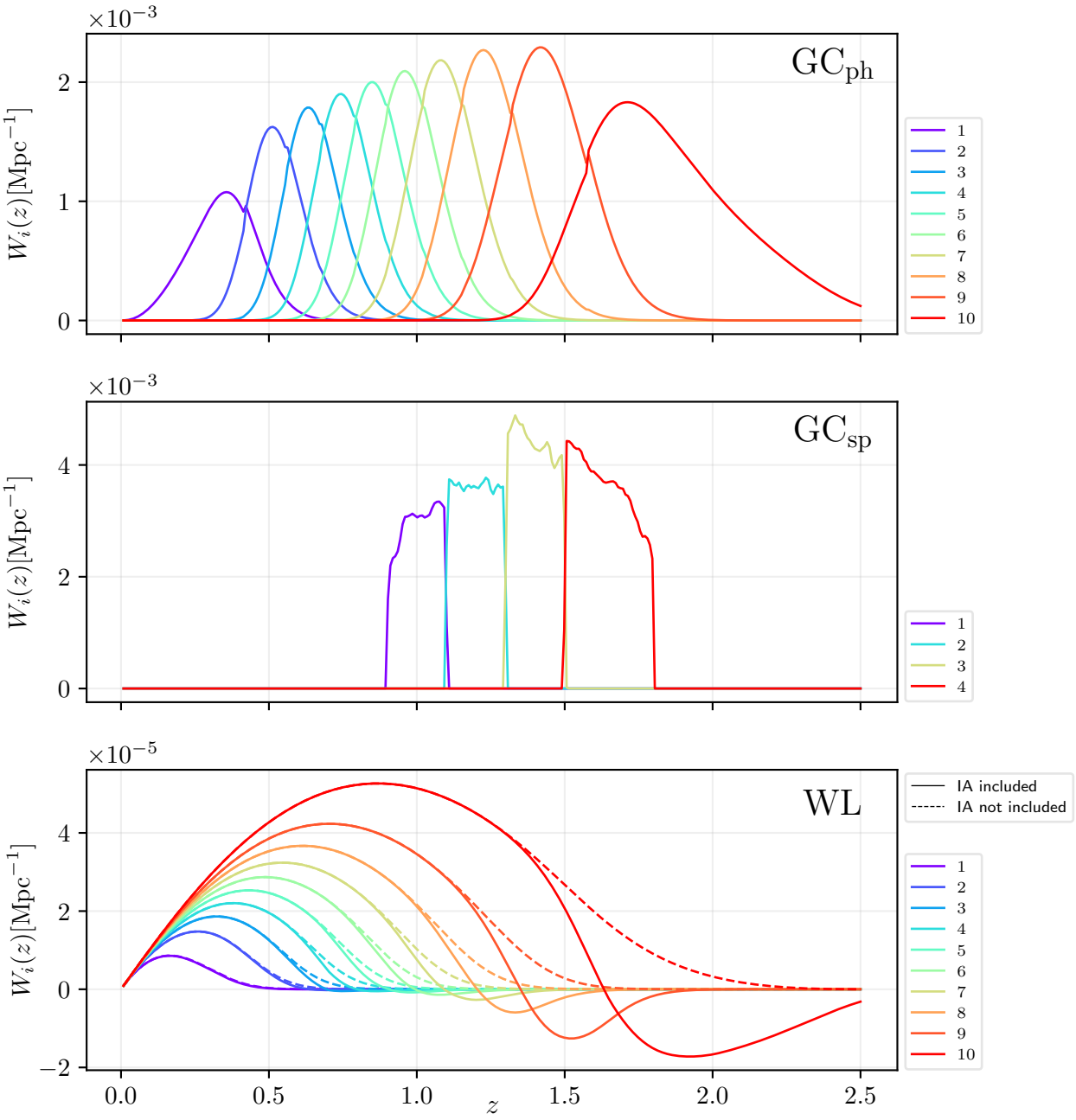

Figure 4.1: Plot of the weight functions for the probes considered in the forecast. Note that for WL at high redshift bins the weight function Eq. (4.21) (solid line) becomes negative, due to the contribution of intrinsic alignment (IA). The shear weight function Eq. (4.20) (dashed line) remains instead always positive as it should be.

probe	c_b	z_b	σ_b	c_o	z_o	σ_o	f_out
$\mathrm{GC_{ph}}$	1.0	0.0	0.050	1.0	0.1	0.05	0.1
$\mathrm{GC_{sp}}$	1.0	0.0	0.001	–	–	–	0.0

Table 4.2: Values of the parameters adopted for the probability distributions defined in Eq. (4.6) and Eq. (4.13). For the $\mathrm{GC_{sp}}$ probe, the outlier parameters are not reported, being the outliers fraction $f_\mathrm{out} = 0$.

4.1.1 Photometric galaxy clustering

In the case of galaxy clustering the starting point to write down the weight function is the galaxy density contrast $\delta^\mathrm{g}(\mathbf{x}, z)$, where $\mathbf{x} = \chi(z)\hat{\mathbf{n}}$ is the comoving distance vector. This is also referred to as the 3D density constrast, since it is a function of both the direction and the distance. As galaxies are tracers of the dark matter distribution, δ^g is expressed in terms of the matter density contrast δ through the galaxy bias b. In Fourier space the galaxy bias can be implicitly defined by the following relation

$$\delta^\mathrm{g}(\mathbf{k}, z) = b(z)\delta(\mathbf{k}, z)\,, \tag{4.7}$$

with $\mathbf{k}$ being the wavevector and k its module. In general the galaxy bias depends also on the scale k, but in this work this k-dependence has been neglected, following the treatment of [5].

The field to be expanded in spherical harmonics for galaxy clustering is $\Delta_i^A(\hat{\mathbf{n}})$, the galaxy density contrast *projected* along the observation direction $\hat{\mathbf{n}}$

$$\Delta_i^A(\hat{\mathbf{n}}) = \int_{z_\mathrm{min}}^{z_\mathrm{max}} \mathrm{d}z\, W_i^A(z) \int \frac{\mathrm{d}^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\hat{\mathbf{n}}\chi(z)} \delta_i^\mathrm{g}(\mathbf{k}, z)\,. \tag{4.8}$$

Here the index A stands for $\mathrm{GC_{ph}}$ or $\mathrm{GC_{sp}}$, and $W_i^A(z)$ is the galaxy clustering weight function

$$W_i^A(z) = b^A(z)\frac{H(z)}{c} n_i^A(z)\,, \tag{4.9}$$

where $n_i^A(z)$ is defined by Eq. (4.5). The true redshift distribution of the photometric sample is modelled as in the *Euclid* redbook [46]

$$\frac{\mathrm{d}N^\mathrm{ph}}{\mathrm{d}z\mathrm{d}\Omega}(z) = N_0^\mathrm{ph} \left(\frac{z}{z_0}\right)^2 \exp\left[-\left(\frac{z}{z_0}\right)^{3/2}\right], \tag{4.10}$$

where $z_0 = 0.9/\sqrt{2}$ and the normalisation factor N_0^ph is chosen such that the surface density of galaxies is equal to 30 galaxies per arcmin^2, corresponding to an expected total number of galaxies of about 1.6×10^9. This redshift distribution is then convolved with the photometric probability distribution $p_\mathrm{ph}(z_\mathrm{p}|z)$ defined in Eq. (4.6).

The redshift evolution of the photometric galaxy bias $b^A(z)$ is modelled as in [5]: for $\mathrm{GC_{ph}}$ the following piecewise constant function is employed

$$b^\mathrm{ph}(z) = b_i^\mathrm{ph} = \sqrt{1 + \bar{z}_i}, \qquad z_i^- < z < z_i^+\,, \tag{4.11}$$

z_i^-	0.001	0.418	0.560	0.678	0.789	0.900	1.019	1.155	1.324	1.576
z_i^+	0.418	0.560	0.678	0.789	0.900	1.019	1.155	1.324	1.576	2.500
$b_i^{\rm ph}$	1.100	1.220	1.272	1.317	1.358	1.400	1.445	1.497	1.565	1.743

Table 4.3: Tomographic bin edges for photometric ($\mathrm{GC_{ph}}$) galaxy clustering, along with the values of the photometric galaxy bias $b_i^{\rm ph}$.

where $\bar{z}_i = (z_i^- + z_i^+)/2$ is the mean redshift of the bin i. As in [5], this forecast employs 10 tomographic bins chosen to be equally populated, i.e. with equal number of galaxies per bin, according to the distribution Eq. (4.10). The bin edges and the photometric galaxy bias values are summarised in Tab. 4.3. The $\mathrm{GC_{ph}}$ weight function is explicitly given by Eq. (4.9) in which $A = \mathrm{ph}$

$$W_i^{\rm ph}(z) = b^{\rm ph}(z)\frac{H(z)}{c}n_i^{\rm ph}(z)\,. \tag{4.12}$$

A plot of this function for all the photometric redshift bins is shown in the upper panel of Fig. 4.1.

4.1.2 Spectroscopic galaxy clustering: 2D approach in the harmonic domain

For the spectroscopic sample the true redshift distribution is the one obtained from the analysis of the *Euclid* SPV2 [74]. The range of the expected distribution of Hα-emitting galaxies is determined by the sensitivity range of NISP detectors. In particular, for a galaxy to be observed by NISP its redshift must fall in the range $z \in [0.9, 1.8]$ (see Chapter 2). The total number of galaxies in the sample is about 2×10^7, which is smaller by a factor ~ 80 than the total number photometric galaxies. In the present work, for numerical reasons[2] it has been decided to convolve the spectroscopic redshift distribution with the probability $p_{\rm sp}(z_{\rm p}|z)$. For this probability is again assumed the functional form Eq. (4.6), but using parameters different from the case of photometric clustering

$$p_{\rm sp}\left(z_{\rm p}|z\right) = \frac{1}{\sqrt{2\pi}\sigma_{\rm b}(1+z)}\exp\left\{-\frac{1}{2}\left[\frac{z - c_{\rm b}z_{\rm p} - z_{\rm b}}{\sigma_{\rm b}(1+z)}\right]^2\right\}. \tag{4.13}$$

In particular, the redshift uncertainty $\sigma_{\rm b}$ is set to 0.001 as specified in the scientific requirements in [46] for *Euclid*, and the fraction of outliers $f_{\rm out}$ is set to zero. The other parameters in Eq. (4.13) are instead assumed to be the same as for $\mathrm{GC_{ph}}$, as summarised in Tab. 4.2.

The spectroscopic galaxy bias $b^{\rm sp}(z)$ also is taken to be a piece-wise constant function, and the values $b_i^{\rm sp}$ in the bins are summarised in Tab. 4.4 along with the bin edges. Both the bias values and the bins are the same that have been used in [5], but in this work also a finer binning has been considered, in particular 12, 24 and 40 equally spaced bins in the range $0.9 < z < 1.8$. The values of the bias for finer binning are obtained by linear interpolation of the values listed in Tab. 4.4.

[2]This operation is aimed at mitigating the sharpness of the boundaries of the spectroscopic bins, to prevent potential numerical instabilities in the computation of the redshift integrals.

z_i^-	0.90	1.10	1.30	1.50
z_i^+	1.10	1.30	1.50	1.80
$b_i^{\rm sp}$	1.46	1.61	1.75	1.90

Table 4.4: Tomographic binning and galaxy bias values $b_i^{\rm sp}$ for spectroscopic $\mathrm{GC_{sp}}$ galaxy clustering.

The $\mathrm{GC_{sp}}$ weight function is simply given by Eq. (4.9) with $A = \mathrm{sp}$

$$W_i^{\rm sp}(z) = b^{\rm sp}(z)\frac{H(z)}{c}n_i^{\rm sp}(z)\,. \tag{4.14}$$

A plot of this function for the case of 4 spectroscopic redshift bins is shown in the middle panel of Fig. 4.1.

4.1.3 Spectroscopic galaxy clustering: 3D approach in Fourier space

In [5] the spectroscopic galaxy clustering is treated by using the Fourier galaxy power spectrum $P_{\rm gg}(k,\mu;z)$, where k is the modulus of the wave mode in Fourier space, and $\mu = \cos\theta = \boldsymbol{k}\cdot\hat{\mathbf{r}}/k$ is the cosine of the angle θ between the wave-vector $\boldsymbol{k}$ and the line-of-sight direction $\hat{\mathbf{r}}$. This section gives an overview of the Fisher matrix computation in the case of the 3D $\mathrm{GC_{sp}}$. For further details refer to Sec. 3.2 in [5].

The main advantage of the Fourier power spectrum approach is that it significantly increases the constraining power of $\mathrm{GC_{sp}}$ as a single probe with respect to the harmonic approach. This is because it allows us to include in a natural way a number of physical effects that carry cosmological information, e.g. the so-called *redshift space distortions* (RSD), which are instead neglected when treating $\mathrm{GC_{sp}}$ in the harmonic domain. The full non-linear model for the observed power spectrum of the Hα galaxies is given by the following equation:

$$P_{\rm obs}\left(k_{\rm ref},\mu_{\rm ref};z\right) = \frac{1}{q_\perp^2 q_\parallel}\left\{\frac{\left[b\sigma_8(z)+f\sigma_8(z)\mu^2\right]^2}{1+\left[f(z)k\mu\sigma_{\rm p}(z)\right]^2}\right\}\frac{P_{\rm dw}(k,\mu;z)}{\sigma_8^2(z)}F_z(k,\mu;z)+P_{\rm s}(z) \tag{4.15}$$

The main physical effects included in this formula are:

- Anisotropies due to RSD: the peculiar velocity contributions to the observed redshifts induce anisotropies in the observed galaxy power spectrum. This effect is encapsulated in the factor enclosed in curly braces in Eq. (4.15), where the denominator accounts for the so-called finger-of-God (FoG) effect by mean of the non-linear parameter $\sigma_{\rm p}(z)$.

- Redshift uncertainty: despite being much more precise than the photometric redshift measurements, the spectroscopic redshifts are still affected by an uncertainty. This smears out the galaxy field along the line of sight, causing a partial suppression of the correlation between galaxy positions. This is included in the multiplicative term $F_z(k,\mu;z) = \mathrm{e}^{-k^2\mu^2\sigma_r^2(z)}$.

- Residual shot noise: this effect is modelled by the additive term $P_{\rm s}(z)$, and it accounts for the non-Poissonian component of the shot noise.

- Alcock-Paczynski (AP) projection effects: in order to convert redshifts into distances for the measurement of the galaxy power spectrum, it is necessary to assume a reference cosmology. Choosing a reference cosmology which is not the true one results in a re-scaling of the components of the wave-vector $\boldsymbol{k}$ in the directions perpendicular and parallel to the line-of-sight. This is known as the Alcock-Paczynski effect, which in Eq. (4.15) is given by the multiplicative factor $1/(q_\perp^2 q_\parallel)$, being $q_\perp$ and $q_\parallel$ the scaling factors of the transverse and radial components of the wave-vector respectively. The subscript "ref" for k and μ in the left hand side of the equation means that k and μ are considered in the reference cosmology.

- Non-linear damping of BAO: another non-linearity effect that is taken into account is the damping of the BAO wiggles in the power spectrum, due to the late-time non-linearities. This is accounted for by using the so-called "de-wiggled" power spectrum $P_{\rm dw}(k,\mu;z)$, which encapsulates an exponential damping factor of the BAO feature.

The Fisher matrix coming from the galaxy Fourier power spectrum $P_{\rm obs}$ is computed by summing up the contributions of the considered redshift bins. The Fisher matrix element for a bin centred at z_i is

$$F_{\alpha\beta}^{\rm bin}\left(z_i\right) = \frac{1}{8\pi^2}\int_{-1}^{1}\mathrm{d}\mu\int_{k_{\rm min}}^{k_{\rm max}}k^2\,\mathrm{d}k\left[\frac{\partial\ln P_{\rm obs}\left(k,\mu;z_i\right)}{\partial\alpha}\frac{\partial\ln P_{\rm obs}\left(k,\mu;z_i\right)}{\partial\beta}\right]V_{\rm eff}\left(z_i;k,\mu\right)\,, \tag{4.16}$$

where $V_{\rm eff}$ is the effective volume of the survey, defined as

$$V_{\rm eff}(k,\mu;z) = V_{\rm s}(z)\left[\frac{n(z)P_{\rm obs}(k,\mu;z)}{n(z)P_{\rm obs}(k,\mu;z)+1}\right]^2\,, \tag{4.17}$$

and $n(z)$ is the number density of Hα galaxies in each redshift bin

$$n(z) = \frac{\mathrm{d}N^{\rm sp}(z)}{\mathrm{d}\Omega\mathrm{d}z}\frac{A_{\rm survey}}{V_{\rm s}(z)}\Delta z\,. \tag{4.18}$$

Finally the Fisher matrix is calculated by summing over the redshift bins:

$$F_{\alpha\beta} = \sum_{i=1}^{N_{z_{\rm bin}}}F_{\alpha\beta}^{\rm bin}\left(z_i\right). \tag{4.19}$$

4.1.4 Weak lensing

As it has already been introduced in Sec. 1.4.4, the gravitational field of large-scale cosmic structure deflects the path of light rays emitted by distant galaxies, distorting the images of the galaxies detected by the observers [75–77]. At the linear level these distortions can be locally decomposed into convergence (κ) and a complex shear γ, which are respectively related to the magnification and shape distortion of the image. In this forecast, only the shear signal γ is considered, and treated in the harmonic domain with the Limber approximation Eq. (4.4). The corresponding weight function for cosmic shear is [5]

$$W_i^{\gamma}(z) = \frac{3}{2}\left(\frac{H_0}{c}\right)^2\Omega_{\rm m,0}(1+z)\chi(z)\int_z^{z_{\rm max}}\mathrm{d}z'\,n_i^{\rm ph}(z')\left[1-\frac{\chi(z)}{\chi(z')}\right]. \tag{4.20}$$

Physically speaking, the integral manifests that weak lensing is a cumulative effect, influenced by all the matter along the line of sight up to the considered redshift z.

The forecast also includes the *intrinsic alignment* (IA), which is the main systematic effect altering weak lensing measurements [78–80]. This is a physical phenomenon due to alignments of nearby galaxies caused by tidal interactions which occur during galaxy formation and evolution. It produces spurious correlations which affect the two-point correlation functions. The correlations caused by IA are indistinguishable from the ones produced by cosmic shear. Therefore, it is necessary to parameterise the IA effect on the lensing angular power spectrum. A possible way to do this is by using the extended non-linear alignment model [5, 81] which consists in the following modification of the cosmic shear weight function

$$W_i^{\rm wl}(z) = W_i^{\gamma}(z) - \mathcal{A}_{\rm IA} C_{\rm IA} \Omega_{\rm m,0} \frac{H(z)\mathcal{F}_{\rm IA}(z)}{cD(z)} n_i^{\rm ph}(z)\,. \tag{4.21}$$

In the above equation, $D(z)$ is the growth factor introduced in Eq. (1.109), which can be directly computed from the linear power spectrum Eq. (1.112). The function $\mathcal{F}_{\rm IA}$ is defined as

$$\mathcal{F}_{\rm IA}(z) \equiv (1+z)^{\eta_{\rm IA}} \left[\frac{\langle L\rangle(z)}{L_\star(z)}\right]^{\beta_{\rm IA}}\,, \tag{4.22}$$

where $\langle L\rangle(z)$ and $L_\star(z)$ are the mean and characteristic luminosity functions, respectively. The intrinsic alignment parameters $\mathcal{A}_{\rm IA}$, $\eta_{\rm IA}$, and $\beta_{\rm IA}$ are treated as nuisance parameters in the analysis. The reference values for the IA parameters are the same adopted in [5]

$$\{\mathcal{A}_{\rm IA}, \eta_{\rm IA}, \beta_{\rm IA}\} = \{1.72, -0.41, 2.17\} \tag{4.23}$$

The parameter $C_{\rm IA} = 0.0134$ is instead kept fixed, since it is completely degenerate with $\mathcal{A}_{\rm IA}$. The scaled luminosity function $\langle L\rangle(z)/L_\star(z)$ is obtained via interpolation of its sampled values available at [73].

4.2 Fisher information matrix

The Fisher information matrix is defined as the expectation value of the hessian matrix of the log-likelihood:

$$F_{\alpha\beta} = -\left\langle \frac{\partial^2 \ln L}{\partial\theta_\alpha \partial\theta_\beta} \right\rangle\,, \tag{4.24}$$

where α and β are the parameter indices. The expected covariance matrix is the inverse of the Fisher matrix:

$$\mathcal{C}_{\alpha\beta} = \left(F^{-1}\right)_{\alpha\beta}\,. \tag{4.25}$$

The diagonal elements of the covariance matrix are the squares of the marginalised 1σ uncertainties on the parameters, where σ is the Gaussian standard deviation:

$$\sigma_\alpha = \sqrt{\mathcal{C}_{\alpha\alpha}}\,. \tag{4.26}$$

The performances are quantified using the FoM mentioned in Sec. 2.1, which has been defined as in [5]

$$\text{FoM} \equiv \sqrt{\det\left(\tilde{F}_{w_0 w_a}\right)} = \left[\sigma^2_{w_0}\sigma^2_{w_a} - \mathcal{C}^2_{w_0 w_a}\right]^{-1/2}\,. \tag{4.27}$$

The notation $\tilde{F}_{w_0 w_a}$ denotes the Fisher information matrix relative to the dark energy equation of state parameters w_0 and w_a, marginalised over all the other free parameters. Note that the Cauchy-Schwarz inequality $\mathcal{C}_{w_0 w_a} \leq \sigma_{w_0}\sigma_{w_a}$ guarantees that the argument of the square root is always greater than or equal to zero. The value of the FoM can be boosted by small errors on w_0 and w_a and/or by their cross-covariance $\mathcal{C}_{w_0 w_a}$. The FoM is inversely proportional to the area of the 1–σ marginalised contour ellipse in the w_0-w_a plane. Tighter constraints on w_0 and w_a lead to smaller ellipses, which in turn means higher FoMs.

4.2.1 Fisher matrix from harmonic approach

Under the assumption that the $a^A_{\ell m}$ of the observed fields follow a gaussian distribution, it can be shown – see Appendix A – that the corresponding Fisher matrix can be computed analytically. In this case the covariance matrix for the harmonic power spectra is

$$\mathrm{Cov}\left[C^{AB}_{ij}(\ell), C^{CD}_{km}(\ell')\right] = \frac{\delta^{\mathrm{K}}_{\ell\ell'}}{2\ell+1}\left[\Sigma^{AC}_{ik}(\ell)\Sigma^{BD}_{jm}(\ell') + \Sigma^{AD}_{im}(\ell)\Sigma^{BC}_{jk}(\ell')\right] , \tag{4.28}$$

with $\delta^{\mathrm{K}}_{\ell\ell'}$ the Kronecker delta symbol, and

$$\Sigma^{AB}_{ij}(\ell) = \frac{1}{\sqrt{f_{\mathrm{sky}}\Delta\ell}}\left[C^{AB}_{ij}(\ell) + N^{AB}_{ij}(\ell)\right] . \tag{4.29}$$

The quantity $N^{AB}_{ij}(\ell)$ is the Poisson shot noise described later in Sec. 4.2.2, f_{sky} is the sky fraction covered by the survey, and $\Delta\ell$ is the spacing between the multipoles in which the $C(\ell)$ are sampled.

In order to compute the Fisher matrix elements it is necessary to define a *data-vector*, which contains the independent values of the $C(\ell)$'s to be included in the computation. In practice, since the tomographic angular power spectra are matrices, they need to be *vectorised*[3] accordingly to their symmetry properties. For example the $\mathrm{GC_{ph}}$ auto-correlation $C^{\mathrm{phph}}_{ij}(\ell)$ at fixed multipole ℓ is a 10×10 symmetric matrix with $10(10+1)/2 = 55$ independent components. At fixed ℓ the data vector can be written in general as

$$\mathbf{C}(\ell) = \{\mathbf{C}_1(\ell), \mathbf{C}_2(\ell), \ldots, \mathbf{C}_{\mathcal{N}}(\ell)\} , \tag{4.30}$$

where $\mathcal{N}$ is the number of the different types of $\mathbf{C}(\ell)$ according to the different probes considered in the analysis. When considering 3 probes at most – WL, $\mathrm{GC_{ph}}$, and $\mathrm{GC_{sp}}$ – the maximum number of angular power spectra that can be computed is $\mathcal{N}_{\mathrm{max}} = 6$. This is because there are 3 possible auto-correlations and 3 possible cross-correlations:

- auto-correlations: $\mathbf{C}^{\mathrm{wlwl}}(\ell)$, $\mathbf{C}^{\mathrm{phph}}(\ell)$, $\mathbf{C}^{\mathrm{phph}}(\ell)$;
- cross-correlations: $\mathbf{C}^{\mathrm{wlph}}(\ell)$, $\mathbf{C}^{\mathrm{wlsp}}(\ell)$, $\mathbf{C}^{\mathrm{phsp}}(\ell)$.

The angular power spectra are evaluated in a finite number N_ℓ of multipoles, therefore the full data-vector $\mathcal{D}$ also includes all the values of the power spectra evaluated at all multipoles

$$\boldsymbol{\mathcal{D}} = \{\mathbf{C}(\ell_1), \mathbf{C}(\ell_2), \ldots, \mathbf{C}(\ell_{N_\ell})\} . \tag{4.31}$$

[3] In this context the term "vectorised" refers to matrix vectorisation, which is defined in Sec. A.2.

The covariance matrix associated to this data-vector is block diagonal, since the Kronecker delta of Eq. (4.28) assures that different multipoles are uncorrelated

$$\mathrm{Cov}\left[\boldsymbol{\mathcal{D}},\boldsymbol{\mathcal{D}}\right]=\begin{pmatrix}\mathrm{Cov}\left[\mathbf{C}(\ell_1),\mathbf{C}(\ell_1)\right] & 0 & \dots & 0\\ 0 & \mathrm{Cov}\left[\mathbf{C}(\ell_2),\mathbf{C}(\ell_2)\right] & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \mathrm{Cov}\left[\mathbf{C}(\ell_{N_\ell}),\mathbf{C}(\ell_{N_\ell})\right]\end{pmatrix}. \quad (4.32)$$

The generic diagonal block, $\mathrm{Cov}\left[\mathbf{C}(\ell),\mathbf{C}(\ell)\right]$, accounts for all the correlations between the different kinds of $C(\ell)$'s included in the data vector, and it is itself a block matrix

$$\mathrm{Cov}\left[\mathbf{C}(\ell),\mathbf{C}(\ell)\right]=\begin{pmatrix}\mathrm{Cov}\left[\mathbf{C}_1(\ell),\mathbf{C}_1(\ell)\right] & \mathrm{Cov}\left[\mathbf{C}_1(\ell),\mathbf{C}_2(\ell)\right] & \cdots & \mathrm{Cov}\left[\mathbf{C}_1(\ell),\mathbf{C}_\mathcal{N}(\ell)\right]\\ \mathrm{Cov}\left[\mathbf{C}_2(\ell),\mathbf{C}_1(\ell)\right] & \mathrm{Cov}\left[\mathbf{C}_2(\ell),\mathbf{C}_2(\ell)\right] & \cdots & \mathrm{Cov}\left[\mathbf{C}_2(\ell),\mathbf{C}_\mathcal{N}(\ell)\right]\\ \vdots & \vdots & \ddots & \vdots\\ \mathrm{Cov}\left[\mathbf{C}_\mathcal{N}(\ell),\mathbf{C}_1(\ell)\right] & \mathrm{Cov}\left[\mathbf{C}_\mathcal{N}(\ell),\mathbf{C}_2(\ell)\right] & \cdots & \mathrm{Cov}\left[\mathbf{C}_\mathcal{N}(\ell),\mathbf{C}_\mathcal{N}(\ell)\right]\end{pmatrix}, \quad (4.33)$$

where the single blocks $\mathrm{Cov}\left[\mathbf{C}_i(\ell),\mathbf{C}_j(\ell)\right]$ are computed according to Eq. (4.28). With these definitions, the Fisher matrix element $F_{\alpha\beta}$ can be calculated as

$$F_{\alpha\beta}=\frac{\partial\boldsymbol{\mathcal{D}}}{\partial\theta_\alpha}\,\mathrm{Cov}^{-1}\left[\boldsymbol{\mathcal{D}},\boldsymbol{\mathcal{D}}\right]\frac{\partial\boldsymbol{\mathcal{D}}}{\partial\theta_\beta}=\sum_{\ell=\ell_1}^{\ell_{N_\ell}}F_{\alpha\beta}(\ell)\,, \quad (4.34)$$

where the second equality follows from the block diagonal form of the covariance, and the fisher matrix element[4] at fixed ℓ is given by

$$F_{\alpha\beta}(\ell)=\frac{\partial\mathbf{C}(\ell)}{\partial\theta_\alpha}^{T}\,\mathrm{Cov}\left[\mathbf{C}(\ell),\mathbf{C}(\ell)\right]^{-1}\frac{\partial\mathbf{C}(\ell)}{\partial\theta_\beta}\,. \quad (4.35)$$

For later convenience, here I report the adopted conventions for naming the Fisher matrices that have been computed in this work:

1. the name of a Fisher matrix is representative of the data vector, and it is composed by different labels, which identify the $C(\ell)$'s contained in the data-vector itself.

2. Within the name of a Fisher matrix, the auto-correlation $C^{\mathrm{AA}}(\ell)$ of the probe A is labeled simply as A, while the cross-correlation $C^{\mathrm{AB}}(\ell)$ between the two probes A and B is denoted as XC(A, B).

3. Square brackets are used to delimit the data-vector extent.

4. The pairwise cross-covariances between the $C(\ell)$'s included in a given data-vector are always taken into account in the computation of the corresponding Fisher matrix.

5. The sum of two Fisher matrices $[F_1]$ and $[F_2]$ is simply denoted by $[F_1]+[F_2]$. This simple sum corresponds to combine $[F_1]$ and $[F_2]$ without taking into account the cross-covariances between their data-vectors, i.e. it is an *independent sum*.

[4] In this work the derivatives of the angular power spectra have been computed using the SteM numerical fitting procedure [82].

The simplest case of a Fisher matrix is when only the auto-correlation of a single probe A is considered in the data-vector. For example, when say $A = \mathrm{GC_{ph}}$, the only observable that can be constructed is the auto-correlation of A with itself. In this case the data-vector is simply

$$\mathbf{C}(\ell) = \left\{ \mathbf{C}^{\mathrm{phph}}(\ell) \right\} . \tag{4.36}$$

Assuming 10 tomographic bins for $\mathrm{GC_{ph}}$, the vector $\mathbf{C}^{\mathrm{phph}}(\ell)$ will have $10(10+1)/2 = 55$ elements. Consequently the covariance matrix is made by a single 55×55 block

$$\mathrm{Cov}\left[\mathbf{C}^{\mathrm{phph}}(\ell), \mathbf{C}^{\mathrm{phph}}(\ell) \right] , \tag{4.37}$$

which takes into account the auto-covariance of $\mathrm{GC_{ph}}$ only. In this case, the Fisher matrix element from Eq. (4.35) becomes

$$F_{\alpha\beta}^{[\mathrm{GC_{ph}}]}(\ell) = \left(\frac{\partial \mathbf{C}^{\mathrm{phph}}(\ell)}{\partial \theta_\alpha} \right)^T \mathrm{Cov}\left[\mathbf{C}^{\mathrm{phph}}(\ell), \mathbf{C}^{\mathrm{phph}}(\ell) \right]^{-1} \frac{\partial \mathbf{C}^{\mathrm{phph}}(\ell)}{\partial \theta_\beta} , \tag{4.38}$$

or, in comma notation

$$F_{\alpha\beta}^{[\mathrm{GC_{ph}}]}(\ell) = \left[\mathbf{C}^{\mathrm{phph}}(\ell)_{,\alpha} \right]^T \mathrm{Cov}\left[\mathbf{C}^{\mathrm{phph}}(\ell), \mathbf{C}^{\mathrm{phph}}(\ell) \right]^{-1} \mathbf{C}^{\mathrm{phph}}(\ell)_{,\beta} . \tag{4.39}$$

When considering two or more probes, multiple combinations can be constructed, depending if *cross-covariances* and *cross-correlations* are included or not in the computation. In this work the word "cross-covariance" are indicates the off-diagonal blocks of the covariance matrix Eq. (4.33), i.e. the blocks $\mathrm{Cov}\left[\mathbf{C}_a(\ell), \mathbf{C}_b(\ell)\right]$ with $a \neq b$. The term "cross-correlation" (signal) is instead used to denote the data-vector entry corresponding to the correlation between two probes, which is therefore something carrying cosmological information.

In the following, the three possible combinations that can be constructed using two probes, A and B, are described in detail.

Fisher matrix from two-probe combinations

There are three possible angular power spectra that can be constructed from two probes A and B:

- auto power spectrum of A, $C^{\mathrm{AA}}(\ell)$;
- auto power spectrum of B, $C^{\mathrm{BB}}(\ell)$;
- cross power spectrum between A and B, $C^{\mathrm{AB}}(\ell)$.

As a concrete example, consider $\mathrm{A} = \mathrm{GC_{ph}}$ and $\mathrm{B} = \mathrm{GC_{sp}}$.

Combining $\mathrm{GC_{ph}}$ and $\mathrm{GC_{sp}}$ as independent probes When assuming $\mathrm{GC_{ph}}$ and $\mathrm{GC_{sp}}$ to be independent, the resulting Fisher matrix will be given by the sum of the Fishers of the two single probes. According to the conventions explained above, the resulting Fisher is denoted as $[\mathrm{GC_{ph}}] + [\mathrm{GC_{sp}}]$. This case is equivalent to build a data-vector including the two auto-correlations

$$\mathbf{C}(\ell) = \left\{ \mathbf{C}^{\mathrm{phph}}(\ell), \mathbf{C}^{\mathrm{spsp}}(\ell) \right\} . \tag{4.40}$$

and setting to zero the off-diagonal blocks of the associated covariance matrix

$$\mathrm{Cov}\left[\mathbf{C}(\ell),\mathbf{C}(\ell)\right]=\begin{pmatrix}\mathrm{Cov}\left[\mathbf{C}^{\mathrm{phph}}(\ell),\mathbf{C}^{\mathrm{phph}}(\ell)\right] & 0\\ 0 & \mathrm{Cov}\left[\mathbf{C}^{\mathrm{spsp}}(\ell),\mathbf{C}^{\mathrm{spsp}}(\ell)\right]\end{pmatrix}. \tag{4.41}$$

Since the matrix in Eq. (4.41) is block diagonal, its inverse will be of the same form [83]:

$$\begin{pmatrix}A & 0\\ 0 & B\end{pmatrix}^{-1}=\begin{pmatrix}A^{-1} & 0\\ 0 & B^{-1}\end{pmatrix}. \tag{4.42}$$

Therefore the matrix product entering the Fisher matrix element will be given by

$$\left(\mathbf{C}^{\mathrm{phph}}(\ell)_{,\alpha}\quad\mathbf{C}^{\mathrm{spsp}}(\ell)_{,\alpha}\right)\begin{pmatrix}\mathrm{Cov}\left[\mathbf{C}^{\mathrm{phph}}(\ell),\mathbf{C}^{\mathrm{phph}}(\ell)\right]^{-1} & 0\\ 0 & \mathrm{Cov}\left[\mathbf{C}^{\mathrm{spsp}}(\ell),\mathbf{C}^{\mathrm{spsp}}(\ell)\right]^{-1}\end{pmatrix}\begin{pmatrix}\mathbf{C}^{\mathrm{phph}}(\ell)_{,\beta}\\ \mathbf{C}^{\mathrm{spsp}}(\ell)_{,\beta}\end{pmatrix}, \tag{4.43}$$

and the resulting Fisher matrix element will be

$$\begin{aligned}F_{\alpha\beta}^{[\mathrm{GC_{ph}}]+[\mathrm{GC_{sp}}]}(\ell)&=\left[\mathbf{C}^{\mathrm{phph}}(\ell)_{,\alpha}\right]^{T}\mathrm{Cov}\left[\mathbf{C}^{\mathrm{phph}}(\ell),\mathbf{C}^{\mathrm{phph}}(\ell)\right]^{-1}\mathbf{C}^{\mathrm{phph}}(\ell)_{,\beta}\\&\quad+\left[\mathbf{C}^{\mathrm{spsp}}(\ell)_{,\alpha}\right]^{T}\mathrm{Cov}\left[\mathbf{C}^{\mathrm{spsp}}(\ell),\mathbf{C}^{\mathrm{spsp}}(\ell)\right]^{-1}\mathbf{C}^{\mathrm{spsp}}(\ell)_{,\beta}\\&=F_{\alpha\beta}^{[\mathrm{GC_{ph}}]}(\ell)+F_{\alpha\beta}^{[\mathrm{GC_{sp}}]}(\ell)\,,\end{aligned} \tag{4.44}$$

which is the sum of the fisher elements associated to the single probes.

Combining $\mathrm{GC_{ph}}$ and $\mathrm{GC_{sp}}$ with cross-covariance The Fisher matrix in this case is denoted as $[\mathrm{GC_{ph}}+\mathrm{GC_{sp}}]$. The data-vector is the same as the previous one Eq. (4.40), but the off-diagonal blocks of the covariance matrix are taken into account

$$\mathrm{Cov}\left[\mathbf{C}(\ell),\mathbf{C}(\ell)\right]=\begin{pmatrix}\mathrm{Cov}\left[\mathbf{C}^{\mathrm{phph}}(\ell),\mathbf{C}^{\mathrm{phph}}(\ell)\right] & \mathrm{Cov}\left[\mathbf{C}^{\mathrm{phph}}(\ell),\mathbf{C}^{\mathrm{spsp}}(\ell)\right]\\ \mathrm{Cov}\left[\mathbf{C}^{\mathrm{spsp}}(\ell),\mathbf{C}^{\mathrm{phph}}(\ell)\right] & \mathrm{Cov}\left[\mathbf{C}^{\mathrm{spsp}}(\ell),\mathbf{C}^{\mathrm{spsp}}(\ell)\right]\end{pmatrix}. \tag{4.45}$$

This matrix is not block-diagonal, hence, when inverting it, the blocks will mix with each other. There exist some formulas based on the Schur complement [83] for the inverse of a 2×2 block matrix, but writing it down does not help to enlighten what happens in this case. From an intuitive point of view, the cross-covariance between two observables should worsen the constraints with respect to combining the two probes as independent. This can be understood with the following argument. If two observables exhibit a non-zero cross-covariance, there will be a mutual correlation between the two. In particular, a change in one of the two – for example induced by a variation of the cosmological parameters – will statistically induce a corresponding variation in the other. This in turn means that the two observables will share an amount of cosmological information, and therefore the total information coming from their combination will be *less* than the direct sum of the two pieces of information carried individually by the two of them. More specifically, combining two correlated observables as if they were independent will end up in a double-counting of the same information, leading to an artificial improvement of the constraining power of their combination.

Combining $\mathrm{GC_{ph}}$, $\mathrm{GC_{sp}}$ and their cross-correlation Here both the covariance and the cross-correlation between $\mathrm{GC_{ph}}$ and $\mathrm{GC_{sp}}$ are taken into account. The resulting Fisher matrix is denoted as $[\mathrm{GC_{ph}} + \mathrm{GC_{sp}} + \mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})]$, and the data-vector includes accordingly the maximal set of the available $C(\ell)$'s

$$\mathbf{C}(\ell) = \left\{ \mathbf{C}^{\mathrm{phph}}(\ell), \mathbf{C}^{\mathrm{phsp}}(\ell), \mathbf{C}^{\mathrm{spsp}}(\ell) \right\} . \tag{4.46}$$

The covariance matrix is made of 3×3 blocks

$$\mathrm{Cov}\left[\mathbf{C}(\ell), \mathbf{C}(\ell)\right] = \begin{pmatrix} \mathrm{Cov}\left[\mathbf{C}^{\mathrm{phph}}(\ell), \mathbf{C}^{\mathrm{phph}}(\ell)\right] & \mathrm{Cov}\left[\mathbf{C}^{\mathrm{phph}}(\ell), \mathbf{C}^{\mathrm{phsp}}(\ell)\right] & \mathrm{Cov}\left[\mathbf{C}^{\mathrm{phph}}(\ell), \mathbf{C}^{\mathrm{spsp}}(\ell)\right] \\ \mathrm{Cov}\left[\mathbf{C}^{\mathrm{phsp}}(\ell), \mathbf{C}^{\mathrm{phph}}(\ell)\right] & \mathrm{Cov}\left[\mathbf{C}^{\mathrm{phsp}}(\ell), \mathbf{C}^{\mathrm{phsp}}(\ell)\right] & \mathrm{Cov}\left[\mathbf{C}^{\mathrm{phsp}}(\ell), \mathbf{C}^{\mathrm{spsp}}(\ell)\right] \\ \mathrm{Cov}\left[\mathbf{C}^{\mathrm{spsp}}(\ell), \mathbf{C}^{\mathrm{phph}}(\ell)\right] & \mathrm{Cov}\left[\mathbf{C}^{\mathrm{spsp}}(\ell), \mathbf{C}^{\mathrm{phsp}}(\ell)\right] & \mathrm{Cov}\left[\mathbf{C}^{\mathrm{spsp}}(\ell), \mathbf{C}^{\mathrm{spsp}}(\ell)\right] \end{pmatrix} . \tag{4.47}$$

In this case the new information coming from the cross-correlation is added to the data-vector, and this contribution is expected to tighten the resulting constraints with respect to the uncorrelated sum. In particular, the cross-correlation is itself a function of the cosmological parameters, meaning that its value will be sensitive to a variation of the parameters themselves. In this sense it is said that adding the *cross-correlation signal* is expected to provide more cosmological information, therefore improving the combined constraints. On the other hand, also all the cross-covariances between the $C(\ell)$'s are being considered in this case, and this will tend to worsen the constraints, as explained in the previous paragraph. So there are two concurring effects, and in principle it is not obvious which of them is dominant. The forecasts performed in my PhD Thesis show that the tightest constraints are actually obtained when both the cross-covariance and cross-correlation are included.

4.2.2 Poisson shot noise

The Poisson shot noise [84] has been implemented in a similar way to what has been done in [5]. It is assumed that only the auto-correlation $C(\ell)$'s are affected by a shot noise, which is independent of the multipole and is a tomographic diagonal matrix

$$N_{ij}^{AB}(\ell) = \delta_{AB}^{\mathrm{K}} \delta_{ij}^{\mathrm{K}} N_i^A . \tag{4.48}$$

So the definition of the noise is reduced to define the quantity N_i^A, which represents the shot noise associated to the probe A at tomographic bin i. For photometric and spectroscopic galaxy clustering in the harmonic domain this is simply given by

$$N_i^{\mathrm{ph}} \equiv \frac{1}{\int_{z_i^-}^{z_i^+} \mathrm{d}z \, \frac{\mathrm{d}N^{\mathrm{ph}}}{\mathrm{d}z\mathrm{d}\Omega}} , \qquad\qquad N_i^{\mathrm{sp}} \equiv \frac{1}{\int_{z_i^-}^{z_i^+} \mathrm{d}z \, \frac{\mathrm{d}N^{\mathrm{sp}}}{\mathrm{d}z\mathrm{d}\Omega}} . \tag{4.49}$$

For weak lensing the definition is instead given by the $\mathrm{GC_{ph}}$ shot noise multiplied by the variance σ_ϵ^2 of the intrinsic galaxy ellipticity

$$N_i^{\mathrm{wl}} \equiv \frac{\sigma_\epsilon^2}{\int_{z_i^-}^{z_i^+} \mathrm{d}z \, \frac{\mathrm{d}N^{\mathrm{ph}}}{\mathrm{d}z\mathrm{d}\Omega}} . \tag{4.50}$$

The value assumed for σ_ϵ is 0.3 as in [5].

As the Kronecker delta $\delta^{\rm K}_{AB}$ of Eq. (4.48) states, no shot noise has been considered for the cross-correlation $C(\ell)$'s. It is in fact commonly assumed for the noises of different tracers to be uncorrelated [5]. This approximation is expected to work well for the cross-correlation between weak lensing and galaxy clustering, since these are different tracers of the same underlying dark matter distribution.

In the cross-correlation of $\mathrm{GC_{ph}}$ with $\mathrm{GC_{sp}}$ the tracers are the galaxies for both the probes, so in principle a shot noise on the cross power spectra (cross-noise) should be present. In this work we checked the $\delta^{\rm K}_{\rm phsp}$ approximation of Eq. (4.48) with the following approach. Given that the Poisson shot noise affecting the two-point function comes from the count of galaxy *self-pairs* – see introduction of [84] – the cross-noise will be due to those galaxies which are in both the photometric and in the spectroscopic sample. The scenario with the highest noise is therefore the one in which the spectroscopic sample is a *proper subset* of the photometric one. So, assuming this worst case scenario, a *subtraction* of the spectroscopic galaxy distribution from the photometric one has been performed. After the subtraction there will be no more overlap between the two samples, and their cross-noise becomes zero by construction. It has been checked that the constraints do not change appreciably after the subtraction, with the FoM and the marginalised uncertainties of the free parameters being negligibly affected.

The shot noise affects in a direct way the diagonal covariance blocks corresponding to the auto-correlation power spectra, as it enters in all the factors of Eq. (4.28). This is what happens for example in the case of the auto-covariance of $\mathrm{GC_{sp}}$

$$\begin{aligned}\mathrm{Cov}\left[C^{\rm spsp}_{ij}(\ell), C^{\rm spsp}_{km}(\ell)\right] \propto \left[C^{\rm spsp}_{ik}(\ell)) + N^{\rm spsp}_{ik}(\ell)\right]\left[C^{\rm spsp}_{jm}(\ell)) + N^{\rm spsp}_{jm}(\ell))\right] + \\ \left[C^{\rm spsp}_{im}(\ell)) + N^{\rm spsp}_{im}(\ell)\right]\left[C^{\rm spsp}_{jk}(\ell)) + N^{\rm spsp}_{jk}(\ell))\right] .\end{aligned} \tag{4.51}$$

However, the fact that the auto-correlation power spectra are contaminated by the shot noise indirectly alters also the other blocks of the covariance matrix. For example the diagonal block corresponding to the auto-covariance of $\mathbf{C}^{\rm phsp}(\ell)$ reads

$$\mathrm{Cov}\left[C^{\rm phsp}_{ij}(\ell), C^{\rm phsp}_{km}(\ell)\right] \propto \left[C^{\rm phph}_{ik}(\ell)) + N^{\rm phph}_{ik}(\ell)\right]\left[C^{\rm spsp}_{jm}(\ell)) + N^{\rm spsp}_{jm}(\ell))\right] + C^{\rm phsp}_{im}(\ell)C^{\rm phsp}_{jk}(\ell)\,, \tag{4.52}$$

and therefore also the terms coming from the inclusion of the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ are affected by the shot noise of both $\mathrm{GC_{ph}}$ and $\mathrm{GC_{sp}}$. Moreover, since the number of galaxies in the spectroscopic sample is ~ 80 times smaller than the galaxies in the photometric one, from Eq. (4.49) it is clear that the $\mathrm{GC_{sp}}$ shot noise will be larger than the one associated to $\mathrm{GC_{ph}}$. In order to quantify the impact of the $\mathrm{GC_{sp}}$ noise the forecast has been performed also in an unrealistic alternative scenario, where this noise is artificially reduced as if the number of spectroscopic galaxies were equal to the number of the photometric ones. This is done in practice by introducing an alternative reduced shot noise for $\mathrm{GC_{sp}}$, defined as follows

$$\tilde{N}^{\rm sp}_i \equiv \frac{N^{\rm tot}_{\rm sp}}{N^{\rm tot}_{\rm ph}} \cdot N^{\rm sp}_i\,, \tag{4.53}$$

where $N^{\rm tot}_{\rm sp}$ ($N^{\rm tot}_{\rm ph}$) is the total number of spectroscopic (photometric) galaxies, computed by integrating the galaxy distribution over its redshift range. As it is discussed in Sec. 4.3, it turns out that the results in the harmonic approach dramatically change using this reduced noise.

4.2.3 Fisher matrix in the hybrid 2D-3D approach

As it has been already pointed out, in this work the observables adopted to compute the Fisher matrix are the angular power spectra. These are computed by mean of a projection on the sky of the galaxy density contrast. From a mathematical point of view, in the Limber approximation this translates into the computation of an integral in the radial direction. For weak lensing, this approach is the most natural one, since the deformation of the images of the source galaxies is a consequence of the cumulative effect of all the matter encountered by the light rays along the line of sight.

Treating also galaxy clustering with this approach allows us to compute cross-covariances and cross-correlation between the main probes in a straightforward way. The main disadvantage of treating the clustering signal in the harmonic domain is that the projection on the celestial sphere consists in a integral along the line of sight. Therefore this projection results in a partial loss of the information from the galaxy density distribution along the radial direction. This in turn ends up in a smaller constraining power from the projected galaxy clustering probe. For photometric clustering, $\mathrm{GC_{ph}}$, this loss is limited, since the redshift resolution is already hampered by the large photometric errors, and therefore one does not lose too much information when projecting along the line of sight. Moreover, the constraining power of $\mathrm{GC_{ph}}$ is boosted by the large number of galaxies in the sample, and this is the main strength of the $\mathrm{GC_{ph}}$ probe.

In the spectroscopic sample there are instead fewer galaxies, and their redshifts are known with much better precision, which is the main strength of the spectroscopic clustering probe. Therefore the 2D projection results in a larger loss of constraining power from $\mathrm{GC_{sp}}$. The tomographic technique can be employed in order to partially recover the radial information about the distances of the galaxies.

The most natural approach to treat $\mathrm{GC_{sp}}$ is the one adopted in [5], which has been briefly described in Sec. 4.1.3, and here it is referred as "3D" or P_k. In this method the 3D galaxy power spectrum is used as the observable, which allows us to naturally exploit the high redshift resolution of the spectroscopic sample, thus significantly increasing the constraining power of $\mathrm{GC_{sp}}$ alone. Nonetheless, in this case it is difficult to compute cross-correlations and cross-covariances between $\mathrm{WL}, \mathrm{GC_{ph}}$, and $\mathrm{GC_{sp}}$.

In this forecast we therefore considered two approaches to combine $\mathrm{GC_{sp}}$ with WL and $\mathrm{GC_{ph}}$. The first is to treat all probes with the angular power spectra, as described in the previous section. In the second approach, which is the topic of this section, the Fisher matrix for the spectroscopic auto-correlation is not computed from the angular power spectra. The contribution of $\mathrm{GC_{sp}}$ is considered by adding *a posteriori* the Fisher matrix computed in the 3D approach of [5], referred here as $\mathrm{GC_{sp}}(P_k)$. This matrix has not been recomputed from scratch by implementing the recipe reported in Sec. 4.1.3. Instead, the official results of [5] have been taken as external inputs. The advantage of this method is that it allows us to include the information coming from the cross-correlations between $\mathrm{GC_{sp}}$ and the other probes in the angular domain, preserving at the same time the higher constraining power of the 3D approach for $\mathrm{GC_{sp}}$. The main drawback is that it is an approximation. In fact, when adding the Fisher matrix of $\mathrm{GC_{sp}}(P_k)$ a posteriori, the cross-covariances between the $\mathrm{GC_{sp}}$ auto-correlation and the other observables are neglected.

In order to be more concrete, here it is reported the case of the combination of $\mathrm{GC_{ph}}$ with $\mathrm{GC_{sp}}$. The starting point is to include only the harmonic $\mathrm{GC_{ph}}$ auto-correlation and the

Forecast settings in the harmonic approach			
Multipoles settings	Optimistic	$\mathrm{GC_{ph}}$	$10 \leq \ell \leq 3000$
		$\mathrm{GC_{sp}}$	$10 \leq \ell \leq 3000$
		WL	$10 \leq \ell \leq 5000$
	Pessimistic	$\mathrm{GC_{ph}}$	$10 \leq \ell \leq 750$
		$\mathrm{GC_{sp}}$	$10 \leq \ell \leq 750$
		WL	$10 \leq \ell \leq 1500$
General settings	Baseline	$\mathrm{GC_{ph}}$	all the 10 bins of Tab. 4.3
		$\mathrm{GC_{sp}}$	reduced shot noise Eq. (4.53)
	Alternative	$\mathrm{GC_{ph}}$	only 4 bins of Tab. 4.3 with $z_i^-, z_i^+ \in [0.9, 1.8]$
		$\mathrm{GC_{sp}}$	reduced shot noise Eq. (4.53)

Table 4.5: Summary of the setting scenarios considered in the forecast. For WL the general settings are not reported since no distinction other than the multipole range has been made. The scenario in which the $\mathrm{GC_{sp}}$ shot noise is reduced *is not intended to be realistic*, and has been studied only to probe the theoretical limitations of the harmonic approach.

harmonic $\mathrm{GC_{ph}} \times \mathrm{GC_{sp}}$ cross-correlation in the data-vector

$$\mathbf{C}(\ell) = \left\{ \mathbf{C}^{\mathrm{phph}}(\ell), \mathbf{C}^{\mathrm{phsp}}(\ell) \right\} . \tag{4.54}$$

Consequently, the covariance matrix is built considering only the upper left sector of Eq. (4.47). The Fisher matrix element is computed with the usual formula Eq. (4.34), and we refer to the resulting matrix with the notation $[\mathrm{GC_{ph}} + \mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})]$. Finally, the Fisher matrix of $\mathrm{GC_{sp}}(P_k)$ is simply added via a sum

$$F\left\{[\mathrm{GC_{ph}} + \mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})] + [\mathrm{GC_{sp}}(P_k)]\right\} = F\left\{[\mathrm{GC_{ph}} + \mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})]\right\} + F\left\{[\mathrm{GC_{sp}}(P_k)]\right\} . \tag{4.55}$$

This procedure is equivalent to neglecting the covariance terms between the $\mathrm{GC_{sp}}$ auto-correlation and the other observables, and performing the computation by using a covariance matrix that can be symbolically written as

$$\begin{pmatrix} \mathrm{Cov}\left[\mathbf{C}^{\mathrm{phph}}, \mathbf{C}^{\mathrm{phph}}\right] & \mathrm{Cov}\left[\mathbf{C}^{\mathrm{phph}}, \mathbf{C}^{\mathrm{phsp}}\right] & 0 \\ \mathrm{Cov}\left[\mathbf{C}^{\mathrm{phsp}}, \mathbf{C}^{\mathrm{phph}}\right] & \mathrm{Cov}\left[\mathbf{C}^{\mathrm{phsp}}, \mathbf{C}^{\mathrm{phsp}}\right] & 0 \\ 0 & 0 & \mathrm{Cov}\left[\mathbf{P}_{\mathrm{obs}}, \mathbf{P}_{\mathrm{obs}}\right] \end{pmatrix} . \tag{4.56}$$

The upper left 2×2 sector contains the covariances between the elements of the $C^{\mathrm{phph}}(\ell)$ and $C^{\mathrm{phsp}}(\ell)$ matrices, organised in block-diagonal form for all multipoles as in Eq. (4.32). Analogously, the lower right corner block represents the auto-covariance of the spectroscopic galaxy Fourier power spectrum for all wave-numbers and redshifts considered. The zeroes correspond to the elements containing the unknown covariances between the 2D and 3D power spectra, which are therefore neglected.

4.3 Results

In this section the results of the forecast are presented. As a convention, the reported marginalised 1–σ uncertainties are always fractional, i.e. divided by the values of the parameters in the reference cosmology in Tab. 4.1. The only exception are the contour plots, which are drawn using the original values of the uncertainties computed via Eq. (4.26). In order to quantitatively assess the impact of cross-covariances and cross-correlations on the forecasts, the corresponding percentage variations of the parameter constraints are plotted with vertical bars. This is done both for the FoM and the relative 1–σ uncertainty, using the following formula

$$\delta X\,[\%] \equiv 100 \cdot \frac{X_b - X_a}{X_a}\,, \tag{4.57}$$

where X generically denotes the FoM or the relative uncertainty $\sigma_\theta/\theta_{\rm fid}$. The subscript a denotes the Fisher matrix whose constraints are used as reference, while b is the Fisher matrix under examination, also referred as the *minuend* in the following.

This section is organised as follows. Sec. 4.3.1 reports the different scenarios and settings with which the forecast has been performed. In Sec. 4.3.2 the results from the combination of $\mathrm{GC_{ph}}$ and $\mathrm{GC_{sp}}$ are presented. Also a comparison is made between $\mathrm{GC_{ph}}$ and $\mathrm{GC_{sp}}$ when treated in the harmonic domain. In Sec. 4.3.3 the results are reported for the combination of WL and $\mathrm{GC_{sp}}$, also comparing the impact on the constraints from $\mathrm{XC(WL, GC_{sp})}$ against $\mathrm{XC(WL, GC_{ph})}$. Finally in Sec. 4.3.4 the constraints coming from the combination of the full set of *Euclid* main probes are shown, focusing on the importance of cross-covariances and cross-correlations between $\mathrm{GC_{sp}}$ and $\mathrm{WL, GC_{ph}}$. In this last section the results of the so-called *Euclid* 6×2pt statistics are reported; with this term it is meant the combination of the all possible two-point functions that can be constructed from $\mathrm{WL, GC_{ph}}$, and $\mathrm{GC_{sp}}$. Using the naming convention for the Fisher matrices introduced in Sec. 4.2, the *harmonic* 6×2pt statistics results are explicitly denoted by

$$[\text{(harmonic)}\,6\times2\text{pt}] \equiv [\mathrm{WL} + \mathrm{GC_{ph}} + \mathrm{GC_{sp}} + \mathrm{XC(WL, GC_{ph})} + \mathrm{XC(WL, GC_{sp})} + \mathrm{XC(GC_{ph}, GC_{sp})}]\,. \tag{4.58}$$

Here the word "harmonic" indicates that all the observables are treated using the angular power spectrum formalism. The corresponding notation of the 6×2pt statistics results in the hybrid approach is

$$[\text{(hybrid)}\,6\times2\text{pt}] \equiv [\mathrm{WL} + \mathrm{GC_{ph}} + \mathrm{XC(WL, GC_{ph})} + \mathrm{XC(WL, GC_{sp})} + \mathrm{XC(GC_{ph}, GC_{sp})}] + [\mathrm{GC_{sp}}(P_k)]\,. \tag{4.59}$$

Therefore, in the hybrid 6×2pt approach everything is treated using $C(\ell)$'s except for the spectroscopic auto-correlation, whose Fisher matrix is added a posteriori and computed via the full, anisotropic, and redshift-dependent galaxy power spectrum as described in Sec. 4.1.3.

4.3.1 Settings description

The scenarios considered in this forecast are summarised in Tab. 4.5. Regarding the multipole range, two scenarios have been studied: one *optimistic* and one *pessimistic*. In the optimistic scenario the multipole range for $\mathrm{GC_{ph}}$ and $\mathrm{GC_{sp}}$ (in the harmonic domain) is set to $\ell \in [10, 3000]$, while it is $\ell \in [10, 5000]$ for WL. In the pessimistic scenario instead the multipole range is $\ell \in [10, 750]$ for $\mathrm{GC_{ph}}$, $\mathrm{GC_{sp}}$ (in the harmonic domain), and $\ell \in [10, 1500]$ for WL.

$GC_{ph} \times GC_{sp}$ FoM results				
GC_{sp} bins	Fisher matrix	FoM	ΔFoM	ΔFoM(%)
–	$[GC_{ph}]$	63.12	–	–
4	$[GC_{ph}] + [GC_{sp}(P_k)]$	230.27	–	–
	$[GC_{ph} + XC(GC_{ph}, GC_{sp})] + [GC_{sp}(P_k)]$	234.54	–	–
	$[GC_{ph}] + [GC_{sp}]$	65.69	–	–
	$[GC_{ph} + GC_{sp}]$	63.95	-1.75	-2.66%
	$[GC_{ph} + GC_{sp} + XC(GC_{ph}, GC_{sp})]$	69.63	$+3.94$	$+6.00\%$
12	$[GC_{ph}] + [GC_{sp}]$	72.02	–	–
	$[GC_{ph} + GC_{sp}]$	70.48	-1.55	-2.15%
	$[GC_{ph} + GC_{sp} + XC(GC_{ph}, GC_{sp})]$	79.87	$+7.85$	$+10.90\%$
24	$[GC_{ph}] + [GC_{sp}]$	85.02	–	–
	$[GC_{ph} + GC_{sp}]$	83.88	-1.13	-1.33%
	$[GC_{ph} + GC_{sp} + XC(GC_{ph}, GC_{sp})]$	108.35	$+23.34$	$+27.45\%$
40	$[GC_{ph}] + [GC_{sp}]$	111.22	–	–
	$[GC_{ph} + GC_{sp}]$	110.39	-0.83	-0.75%
	$[GC_{ph} + GC_{sp} + XC(GC_{ph}, GC_{sp})]$	153.71	$+42.48$	$+38.20\%$

Table 4.6: Table reporting the FoM for GC_{ph}, GC_{sp} and their cross-correlation. The ΔFoM column contains the variation of the figure of merit with respect to the independent sum $[GC_{ph}] + [GC_{sp}]$ for the given number of bins. The ΔFoM(%) column is calculated by taking ΔFoM as a percentage of the FoM of $[GC_{ph}] + [GC_{sp}]$.

For galaxy clustering one more assumption has been made, in order to better understand the differences between GC_{ph} and GC_{sp} when both are treated in the harmonic domain. For GC_{sp} the alternative reduced shot noise described in Sec. 4.2.2 has been employed, in order to assess its impact on the results. However it is important to stress that this scenario in which the GC_{sp} shot noise has been reduced *is not meant to be realistic*. It has been considered only to understand the reason for the different constraining power of the GC_{sp} and GC_{ph} probes. For the same reason, in the case of photometric clustering the analysis has been performed using also an alternative tomographic binning, where only the 4 photometric bins entirely comprised in the redshift range $z \in [0.9, 1.8]$ are considered. This has been done in order to make both GC_{ph} and $XC(WL, GC_{ph})$ comparable with GC_{sp} and $XC(WL, GC_{sp})$ respectively. In fact, apart from the shape of the galaxy distribution and the values of the galaxy bias, in the harmonic domain the two main differences between GC_{ph} and GC_{sp} are the shot noise and the redshift range of the galaxy catalogue. As said in Sec. 4.2.2, the shot noise of GC_{sp} is much larger than the GC_{ph} one, since the spectroscopic sample has 80 times fewer galaxies than the photometric one. Moreover, the redshift range covered by the photometric galaxy distribution is $[0.001, 2.5]$, while for the spectroscopic galaxy distribution it is limited to $z \in [0.9, 1.8]$ due to the wavelength range in which NISP is sensitive.

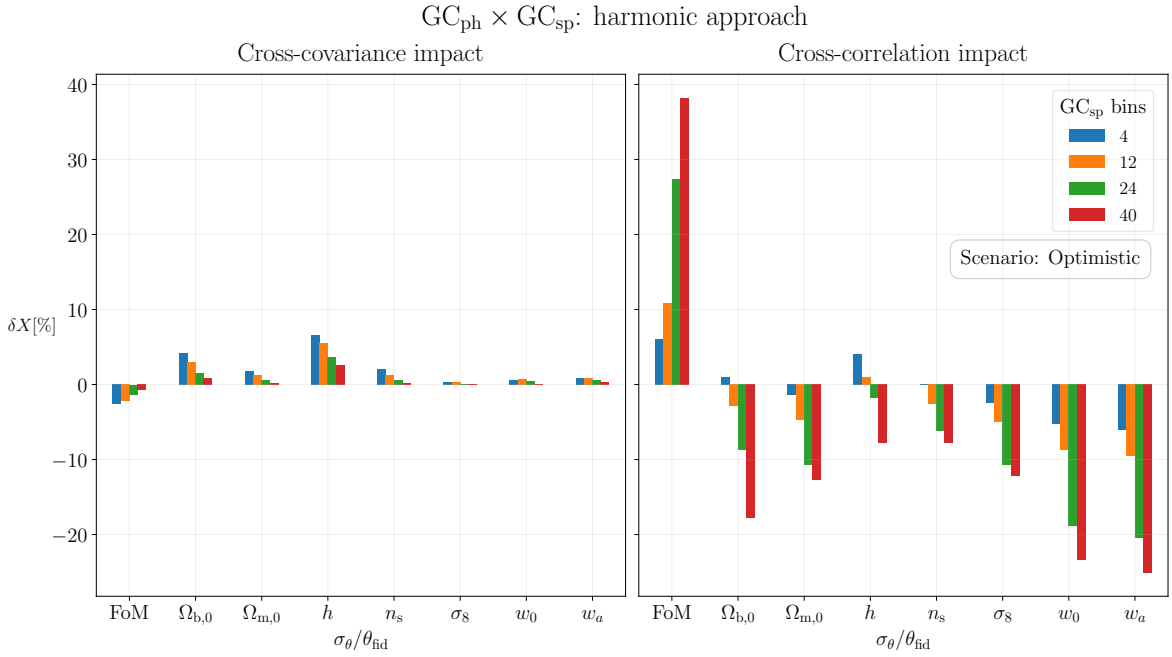

Figure 4.2: Impact, on 1–σ parameter errors and FoM, of cross-covariance and cross-correlation in the harmonic approach for the combination of GC_{ph} and GC_{sp}. The left panel shows the impact of the cross-covariance, quantified with the percentage differences between $[GC_{ph} + GC_{sp}]$ and $[GC_{ph}] + [GC_{sp}]$. In the right panel the impact of the cross-correlation is reported, quantified by percentage differences of $[GC_{ph} + GC_{sp} + XC(GC_{ph}, GC_{sp})]$ with respect to $[GC_{ph}] + [GC_{sp}]$.

4.3.2 Combining photometric and spectroscopic clustering

In this section I report the results for the combination of GC_{ph} and GC_{sp}. As said above, in this case the Fisher matrix in the harmonic approach is denoted as $[GC_{ph} + GC_{sp} + XC(GC_{ph}, GC_{sp})]$, while in the hybrid one as $[GC_{ph} + XC(GC_{ph}, GC_{sp})] + [GC_{sp}(P_k)]$. In the following I discuss the impact on the constraints given by the inclusion of the GC_{ph}-GC_{sp} cross-covariance and the $XC(GC_{ph}, GC_{sp})$ cross-correlation. In this section, I take as reference values for the percentages the constraints coming from the independent combination of GC_{ph} and GC_{sp}, that is $[GC_{ph}] + [GC_{sp}]$ in the harmonic approach and $[GC_{ph}] + [GC_{sp}(P_k)]$ in the hybrid one. The results for this combination are reported in the optimistic scenario only for brevity.

Impact of the GC_{ph}-GC_{sp} cross-covariance on parameter constraints

According to the notation convention defined in Sec. 4.2.1, in the case of the cross-covariance alone, the Fisher matrix $[GC_{ph} + GC_{sp}]$ is computed in the harmonic domain by taking into account only the cross-covariance between GC_{ph} and GC_{sp}. The notation $[GC_{ph}] + [GC_{sp}]$ represents instead the simple sum of the Fisher matrices of the two probes, which does not take the cross-covariance into account. From Tab. 4.6 it can be seen that in any case the cross-covariance slightly worsens the FoM, as expected. Anyway the contribution is always smaller

than 3%, in particular it is -2.66% with 4 bins and -0.75% with 40 bins used for $\mathrm{GC_{sp}}$. This decreasing trend can be understood as follows: the finer the tomographic binning of $\mathrm{GC_{sp}}$ the smaller the support of the integrand of the off-diagonal terms $\mathbf{C}^{\mathrm{phsp}}(\ell)$. Intuitively, the value of the integral over each of the 4 thick bins is diluted into more thinner bins when a finer binning is used. Hence the off-diagonal block Cov $\left[\mathbf{C}^{\mathrm{phph}}(\ell), \mathbf{C}^{\mathrm{spsp}}(\ell)\right]$ and its transposed counterpart become larger and sparser as the number of bins is increased, and therefore the cross-covariance contribution becomes smaller. Physically this could be understood considering that for 40 bins the loss of information due to the projection transversely to the line of sight is less severe than for 4 bins only.

The same trend is observed also for the marginalised uncertainties on the cosmological parameters, as it can be seen by Fig. 4.2, where the percentage difference between the uncertainties, as defined in Eq. (4.57), are drawn as vertical bars. In particular, the impact of the cross-covariance is shown in the left panel as the comparison between $[\mathrm{GC_{ph}} + \mathrm{GC_{sp}}]$ and $[\mathrm{GC_{ph}}] + [\mathrm{GC_{sp}}]$, taking this last as reference. The parameters mostly affected by the covariance are $\Omega_{\mathrm{b},0}$ and h, whose constraints in the 4 bin setting worsen by 4% and 6% respectively. However, as soon as the binning is refined, the effect gets smaller, becoming about 0.9% for $\Omega_{\mathrm{b},0}$ and 2.5% for h. The uncertainties on $\Omega_{\mathrm{m},0}$, n_{s}, and σ_8 are instead affected by less than 2% for all the binning settings. This outcome is confirmed also by Fig. 4.3, which reports the relative marginalised uncertainties and the FoM as horizontal bars. The $[\mathrm{GC_{ph}} + \mathrm{GC_{sp}}]$ bars (in blue) always have practically the same length as the $[\mathrm{GC_{ph}}] + [\mathrm{GC_{sp}}]$ bars (in cyan), and they look more and more similar as the number of $\mathrm{GC_{sp}}$ bins increases.

In conclusion the cross-covariance between $\mathrm{GC_{ph}}$ and $\mathrm{GC_{sp}}$ can be considered negligible, as it does not change the uncertainties on the cosmological parameters by more than 6% and the FoM by more than 3%.

Impact of the XC($\mathbf{GC_{ph}}, \mathbf{GC_{sp}}$) signal on parameter constraints

Harmonic approach In the harmonic approach the contribution to the FoM coming from the XC($\mathrm{GC_{ph}}, \mathrm{GC_{sp}}$) cross-correlation is always positive, as reported in Tab. 4.6, and it is about 6% (38%) with 4 (40) bins for $\mathrm{GC_{sp}}$. This gain is visible also in Fig. 4.3, where the FoM and the marginalised uncertainties from the $[\mathrm{GC_{ph}} + \mathrm{GC_{sp}} + \mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})]$ Fisher matrix are represented with green bars. The improvements on the errors for w_0, w_a, and consequently on the FoM, are particularly visible for finer binnings, coherently with the results reported in the table.

The marginalised uncertainties on the cosmological parameters also improve when including the XC signal, but the gain is more modest than the one on the FoM, as it is shown in the right panel of Fig. 4.2. In fact, the biggest improvements are for w_0 and w_a, whose uncertainties get smaller at most by 23% and 25%, respectively. Instead, the uncertainty on the baryon density parameter, $\Omega_{\mathrm{b},0}$, slightly increases by about 1% with 4 bins and becomes smaller by 18% with 40 bins. The small worsening on this parameter in the case of 4 bins has been attributed to the contribution of the cross-covariances between the $C(\ell)$'s included in the data-vector of the Fisher matrix $[\mathrm{GC_{ph}} + \mathrm{GC_{sp}} + \mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})]$, which are not taken into account in the simple sum $[\mathrm{GC_{ph}}] + [\mathrm{GC_{sp}}]$. It is possible that with 4 bins the gain coming from the inclusion of the XC signal is compensated by the covariance contribution, producing a net (small) worsening. However, as soon as the number of $\mathrm{GC_{sp}}$ bins increases, the positive contribution of the cross-correlation signal starts to dominate, and the constraints on $\Omega_{\mathrm{b},0}$ to

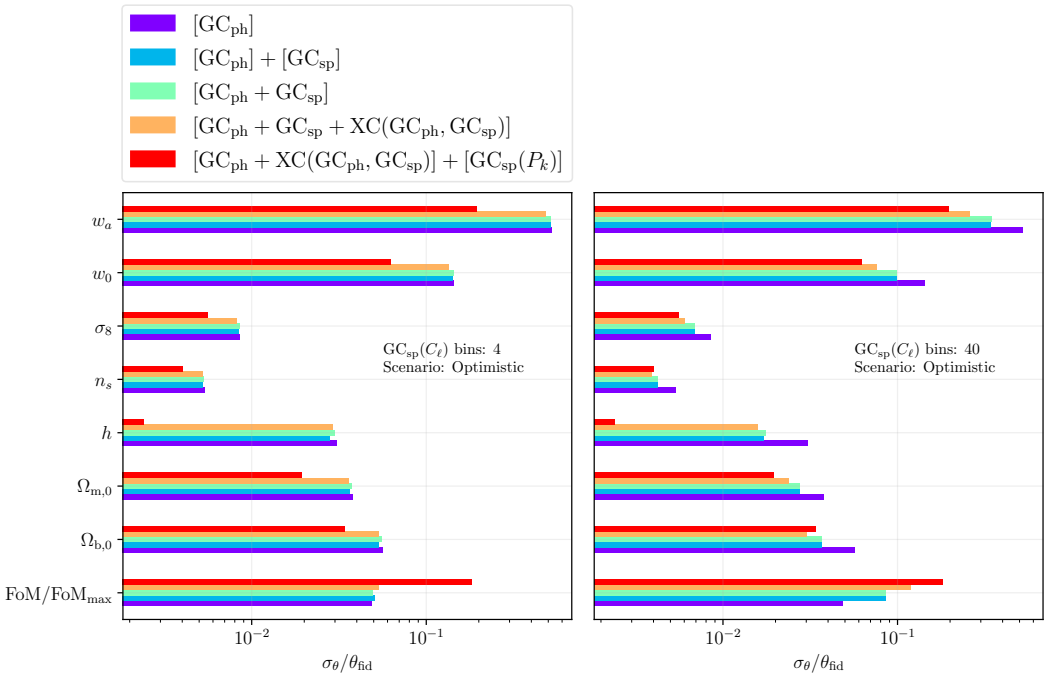

Figure 4.3: Comparison between harmonic and hybrid harmonic-Fourier approaches for the case of $\mathrm{GC_{ph}} \times \mathrm{GC_{sp}}$.

improve too. The uncertainties on the other parameters also improve, in particular $\Omega_{\mathrm{m},0}$ and σ_8 gain at most 12% when 40 bins are used, as well as h and n_s which improve by 7% at maximum.

Hybrid approach In the hybrid approach instead, the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ has quite a small impact on the constraints: the marginalised uncertainties and FoM from the $[\mathrm{GC_{ph}} + \mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})] + [\mathrm{GC_{sp}}(P_k)]$ Fisher matrix are very similar to the ones from $[\mathrm{GC_{ph}}] + [\mathrm{GC_{sp}}(P_k)]$, as it is qualitatively shown in the barplots of Fig. 4.3. A more detailed analysis is reported in Fig. 4.4, where the percentage variations on the constraints resulting from adding the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ are shown. The variation on the FoM is comprised between +1.5% and +2%, so there is a very small improvement when including $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ in the hybrid approach. This is also the case for the marginalised uncertainties which improve less than 1%. Therefore, it can be concluded that, in the hybrid approach, the inclusion of the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ cross-correlation in the combination of $\mathrm{GC_{ph}}$ and $\mathrm{GC_{sp}}$ has a negligible impact on the results.

Hybrid approach vs harmonic approach

As Fig. 4.5 shows, the hybrid approach performs better than the harmonic one in constraining almost all the cosmological parameters, even if the constraining power of the harmonic approach significantly improves with increasing the number of $\mathrm{GC_{sp}}$ tomographic bins. The harmonic approach has a FoM of 69 (153) for 4 (40) spectroscopic bins, while the hybrid one provides a FoM of 234, which is 236% (52%) larger than in the former case. Thus the

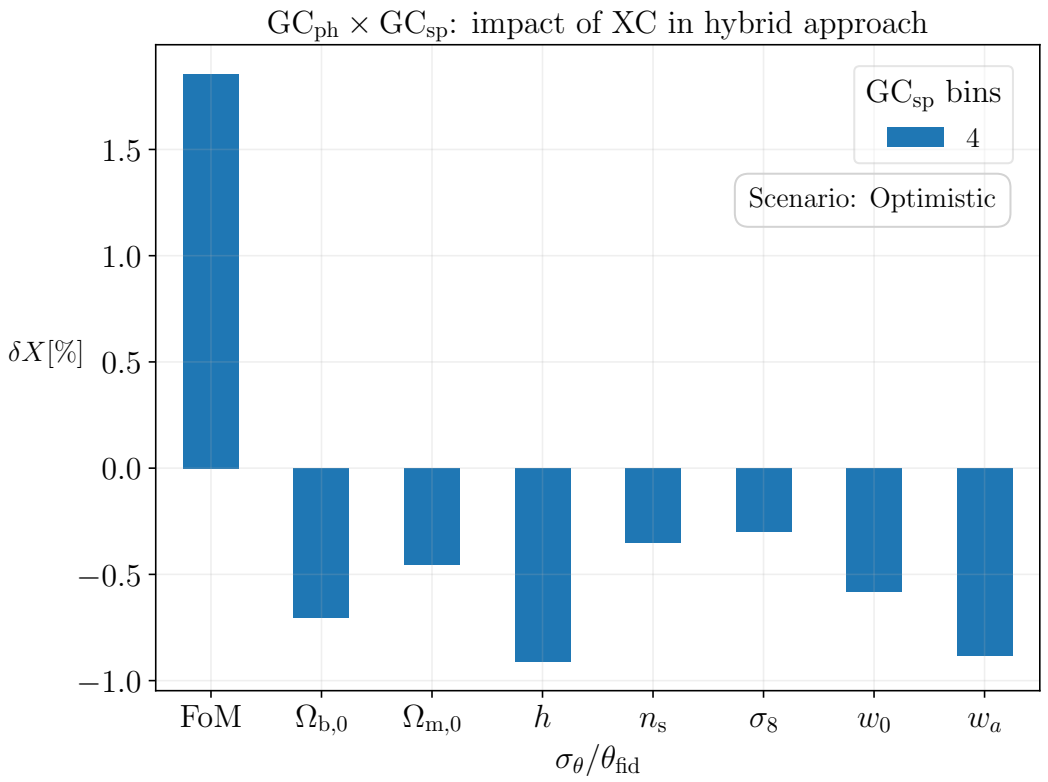

Figure 4.4: Impact, on FoM and marginalised 1–σ errors, of the XC(GC$_{ph}$, GC$_{sp}$) in the hybrid approach, for the optimistic scenario. The reference here are the constraints coming from [GC$_{ph}$] + [GC$_{sp}$(P_k)], while the minuend for the percentage difference is [GC$_{ph}$ + XC(GC$_{ph}$, GC$_{sp}$)] + [GC$_{sp}$(P_k)]. Note that in the hybrid approach only 4 tomographic bins were used to compute the XC(GC$_{ph}$, GC$_{sp}$) angular power spectra, for consistency with the Fourier power spectrum analysis.

hybrid approach is better even when 40 spectroscopic bins are used for the harmonic one. In particular, this is true for the reduced Hubble constant h. In this case the hybrid approach performs remarkably better, with a gain on the marginalised uncertainty which is always between 75% and 100%, depending only slightly on the number of GC$_{sp}$ bins used for the harmonic approach. This is expected, since the hybrid approach takes advantage of the 3D power spectrum as an observable, in which radial BAO and RSD are included. The constraints on the other parameters appear instead quite sensitive to the GC$_{sp}$ binning. The differences on the w_0 and w_a uncertainties significantly decrease from more than 50% with 4 bins to about 20% with 40 bins, as it could be expected from the FoM differences between the hybrid approach and the harmonic one. For $\Omega_{m,0}$, and σ_8 the hybrid approach is still better than the harmonic one, even if the difference between the two approaches decreases significantly with the number of bins too.

The only exceptions are given by the baryon density, $\Omega_{b,0}$, and the spectral index, n_s. In particular, the $\Omega_{b,0}$ uncertainty from the hybrid approach is more than 25% smaller than from the harmonic one with 4 spectroscopic bins, but the situation gets reversed with 40 bins. In

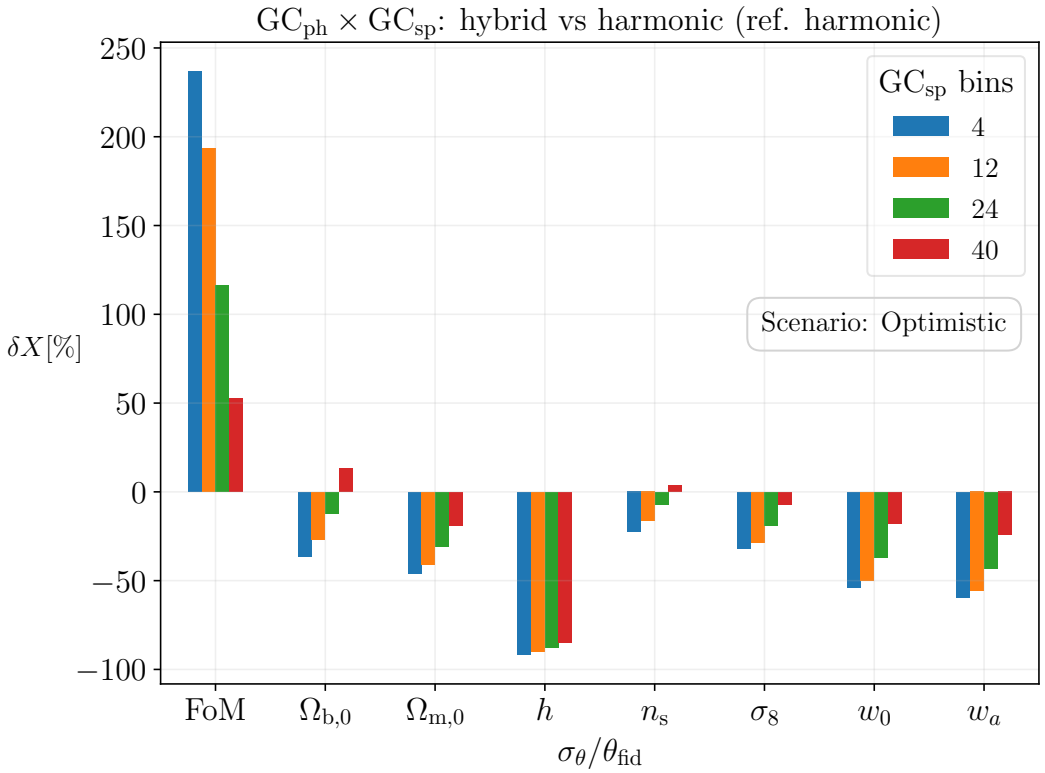

Figure 4.5: Percentage difference between FoM and marginalised 1–σ errors of the hybrid and harmonic approaches, for the combination of GC$_{ph}$ and GC$_{sp}$ in the baseline optimistic scenario. The percentage differences are normalised to the results of the harmonic approach.

the latter case the harmonic approach provides $\sim 10\%$ better constraints on $\Omega_{b,0}$ than the hybrid one. The n_s uncertainty in the harmonic approach with 4 bins is about 20% larger than in the hybrid one. Instead, when 40 bins are used for the harmonic approach, the n_s uncertainty given by the hybrid approach is a few percents larger than the one given by the former.

To conclude, for the combination of GC$_{ph}$ and GC$_{sp}$, the hybrid approach always provides better constraints than the harmonic one. Increasing the number of bins of the harmonic approach improves its performances. Using 40 bins allows us to reach the performances of the hybrid approach for some parameters – n_s and $\Omega_{b,0}$ – but a large gap still remains for the other constraints, especially for h and the FoM.

Comparing photometric and spectroscopic clustering in the harmonic domain

As anticipated in Sec. 4.3.1 two alternative configurations have been considered both for GC$_{ph}$ and GC$_{sp}$. The baseline configuration for GC$_{ph}$ consists in using all the 10 redshift bins reported in Tab. 4.3, while in the alternative configuration only the 4 bins strictly contained in the range $[0.9, 1.8]$ are considered. For spectroscopic clustering instead the baseline settings correspond to use the shot noise as calculated from the second equation of Eq. (4.49), while the alternative consists in using the artificially reduced version Eq. (4.53) of the noise. In fact, from

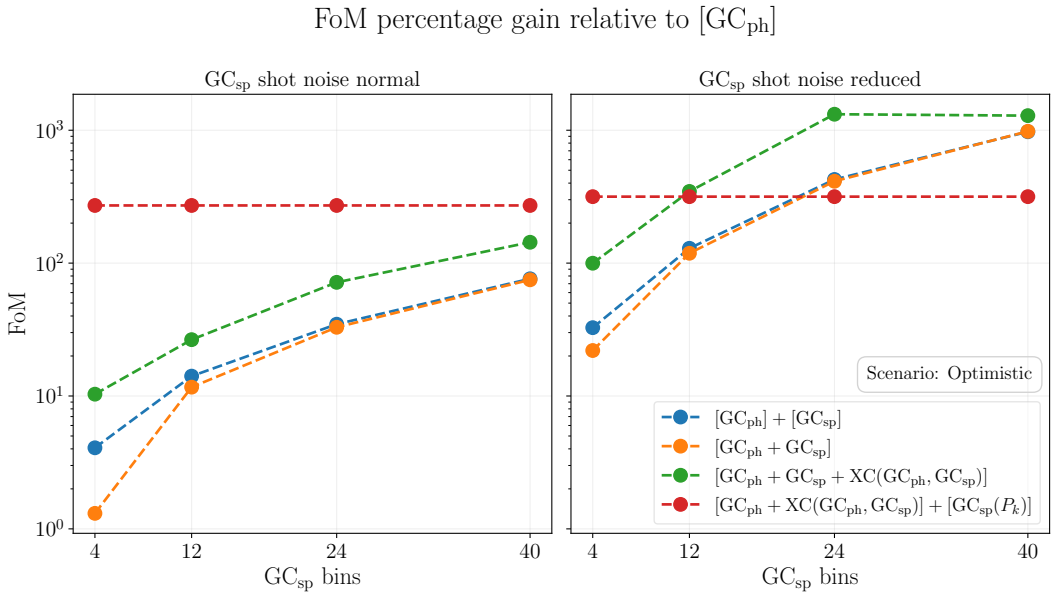

Figure 4.6: Percentage gain on the FoM relative to photometric clustering alone ([GC$_{ph}$]). In the left panel the computation has been done using the standard GC$_{sp}$ shot noise Eq. (4.49), while in the right panel the reduced version Eq. (4.53) has been used.

further investigation it turned out that a great limitation of the harmonic approach is due to the shot noise associated to the GC$_{sp}$ auto-correlation $C(\ell)$'s. This is in fact much higher than the one associated to GC$_{ph}$, since the expected number of Hα-emitting galaxies in the *Euclid* catalogue ($\sim 2 \times 10^7$) is smaller than the expected number of galaxies in the photometric sample ($\sim 1.6 \times 10^9$). Therefore, in order to quantify the impact of the spectroscopic shot noise, the forecast has been performed also in the artificially reduced shot noise setting, as explained in Sec. 4.2.2. The results are shown in Fig. 4.6: in the left panel the standard GC$_{sp}$ shot noise computed from Eq. (4.49) is employed, while in the right one the reduced noise Eq. (4.53) is used. From the panels it is manifest that lowering the shot noise systematically boosts the results of the harmonic approach by about one order of magnitude, making its performance comparable or even better than in the hybrid case. In particular, as it can be expected, the observable which most improves its performance thanks to the noise reduction is the harmonic GC$_{sp}$ auto-correlation. Instead, the XC(GC$_{ph}$, GC$_{sp}$) is not affected as the shot noise does not enter its definition, as can be observed from Eq. (4.48).

Instead, when considering the behaviour of the [GC$_{ph}$ + XC(GC$_{ph}$, GC$_{sp}$)] + [GC$_{sp}$(P_k)] Fisher matrix, as compared to the corresponding one in the harmonic domain, its performance remains stable against the change of the shot noise level, since the possible change of the latter would enter only GC$_{sp}$(P_k), which is a fixed external input. Therefore, while in the harmonic case the survey performance increases with the noise reduction and the increasing of the number of spectroscopic bins, in the hybrid case the performance stay unchanged since both the noise level and the bin number are fixed to the values of the external input.

It is important to stress that the condition of reduced shot noise for GC$_{sp}$ *is not realistic*, since it is realised only if the number of galaxies in the spectroscopic sample equals the number of the photometric one, which is not the case. This check has been done in order to better

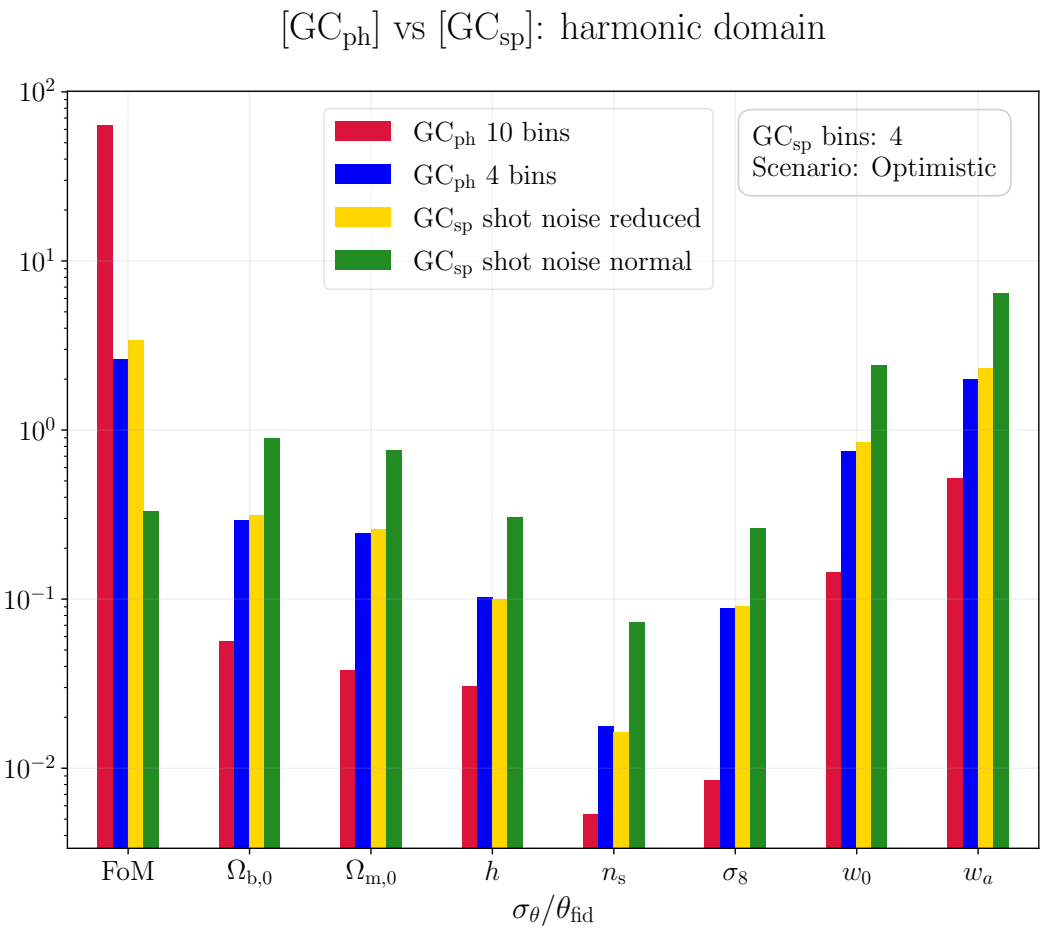

Figure 4.7: Comparison between the constraints coming from [GC$_{ph}$] and [GC$_{sp}$] in the harmonic approach, in the different configurations described in Sec. 4.3.1. The scale on the vertical axis is logarithmic to make both the relative marginalised uncertainties and the FoM visible in the same plot.

understand the limitations of GC$_{sp}$ in the harmonic approach.

The results of considering the same shot noise level and redshift range for GC$_{sp}$ and GC$_{ph}$ are reported in Fig. 4.7. The height of the bars labeled with the cosmological parameters is equal to the relative marginalised uncertainty for that particular parameter. Regarding the FoM the height is instead given by the value of the FoM itself. From the plot it is evident that the constraints coming from GC$_{ph}$ restricted to the 4 bins in the spectroscopic range (blue bars) are very close to the ones of GC$_{sp}$ when the shot noise is reduced (yellow bars). Thus in these special conditions the two probes are essentially equivalent. This is expected, since the functional form of the weight for galaxy clustering is always the same, with the photometric function W_i^{ph} differing from the spectroscopic one W_i^{sp} only for the shape of the galaxy distribution and the values of the galaxy bias.

Fig. 4.7 shows also the two probes in the baseline configuration, in which the 10 bins listed in Tab. 4.3 are used for GC$_{ph}$ (red bars) and the more realistic shot noise Eq. (4.49) is used

WL × GC_{sp} FoM results				
GC_{sp} bins	Fisher matrix	FoM	ΔFoM	ΔFoM(%)
4	$[WL] + [GC_{sp}(P_k)]$	158.13	–	–
	$[WL + XC(WL, GC_{sp})] + [GC_{sp}(P_k)]$	182.74	–	–
	$[WL] + [GC_{sp}]$	74.72	–	–
	$[WL + GC_{sp}]$	74.37	−0.35	−0.47%
	$[WL + GC_{sp} + XC(WL, GC_{sp})]$	103.56	+28.83	+38.59%
12	$[WL] + [GC_{sp}]$	92.90	–	–
	$[WL + GC_{sp}]$	92.77	−0.13	−0.14%
	$[WL + GC_{sp} + XC(WL, GC_{sp})]$	131.38	+38.47	+41.41%
24	$[WL] + [GC_{sp}]$	111.41	–	–
	$[WL + GC_{sp}]$	111.42	+0.01	+0.0021%
	$[WL + GC_{sp} + XC(WL, GC_{sp})]$	155.18	+43.76	+39.28%
40	$[WL] + [GC_{sp}]$	141.12	–	–
	$[WL + GC_{sp}]$	141.17	+0.06	+0.042%
	$[WL + GC_{sp} + XC(WL, GC_{sp})]$	188.45	+47.34	+33.55%

Table 4.7: Table reporting the FoM for WL, GC_{sp} and their cross-correlation The ΔFoM column contains the variation of the figure of merit with respect to the independent sum $[WL] + [GC_{sp}]$ for the given number of bins. The ΔFoM(%) column is calculated by taking ΔFoM as a percentage of the FoM of $[WL] + [GC_{sp}]$.

for GC_{sp} (green bars). On the one hand, the usage of all bins for GC_{ph} makes a big difference, improving the FoM by a factor of ~ 30 and reducing the uncertainties of about one order of magnitude on average. On the other hand, the realistic shot noise of GC_{sp} significantly affects its performances, reducing the FoM by a factor of ~ 10 and making the uncertainties larger.

The conclusion is therefore that the shot noise and the redshift range of the galaxy sample are what makes the difference between GC_{ph} and GC_{sp} in terms of constraining power, when both are treated in the harmonic domain. When GC_{ph} is restricted to the same range of GC_{sp} and the shot noise of the latter is reduced to the same level of the former, their constraints become comparable with each other.

4.3.3 Combining weak lensing and spectroscopic clustering

In this section I report the results from the combination of WL and GC_{sp}. The Fisher matrix in the harmonic approach is $[WL + GC_{sp} + XC(WL, GC_{sp})]$, while in the hybrid one it is $[WL+XC(WL, GC_{sp})]+[GC_{sp}(P_k)]$. I quantify the impact on parameter constraints of the WL-GC_{sp} cross-covariance and the $XC(WL, GC_{sp})$ cross-correlation using percentage differences. I take as reference values for the percentages the constraints coming from the Fisher matrix corresponding to independent combination of WL and GC_{sp}, that is $[WL] + [GC_{sp}]$ in the harmonic approach and $[WL] + [GC_{sp}(P_k)]$ in the hybrid one. For brevity, the results of this combination are reported in the optimistic scenario only.

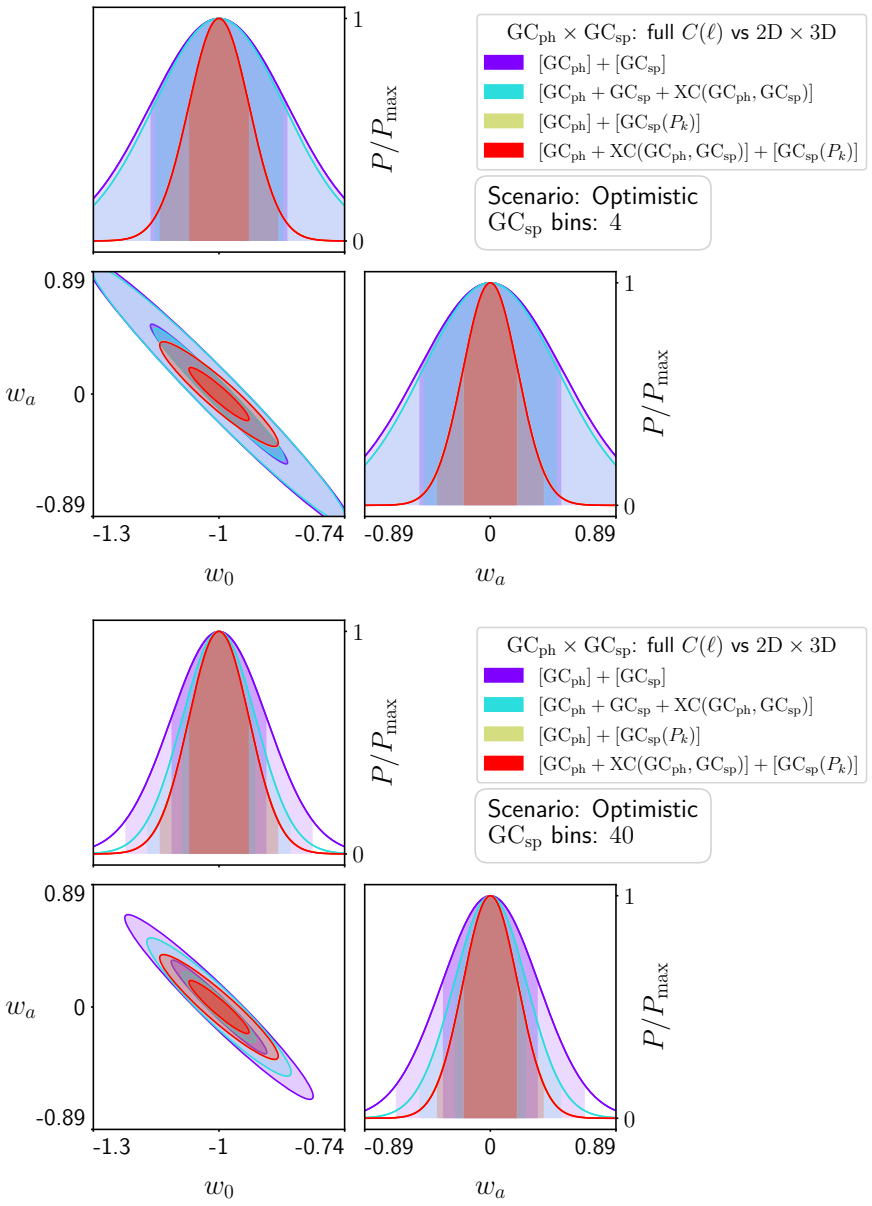

Figure 4.8: Contour plots showing the comparison between the two approaches in the combination of GC_{ph} and GC_{sp}.

Impact of the WL-GC_{sp} cross-covariance on parameter constraints

In Tab. 4.7 the FoM values resulting from the combinations of WL and GC_{sp} are reported. In this case the cross-covariance is even more negligible than in the GC_{ph}-GC_{sp} case, always impacting by less than 0.5% on the figure of merit. This effect gets smaller as the number of bins increases, starting from -0.47% with 4 spectroscopic bins up to $+0.042\%$ with 40 spectroscopic bins respectively. This last is unexpectedly positive, but the variation is so small that can be considered as a numerical fluctuation around zero.

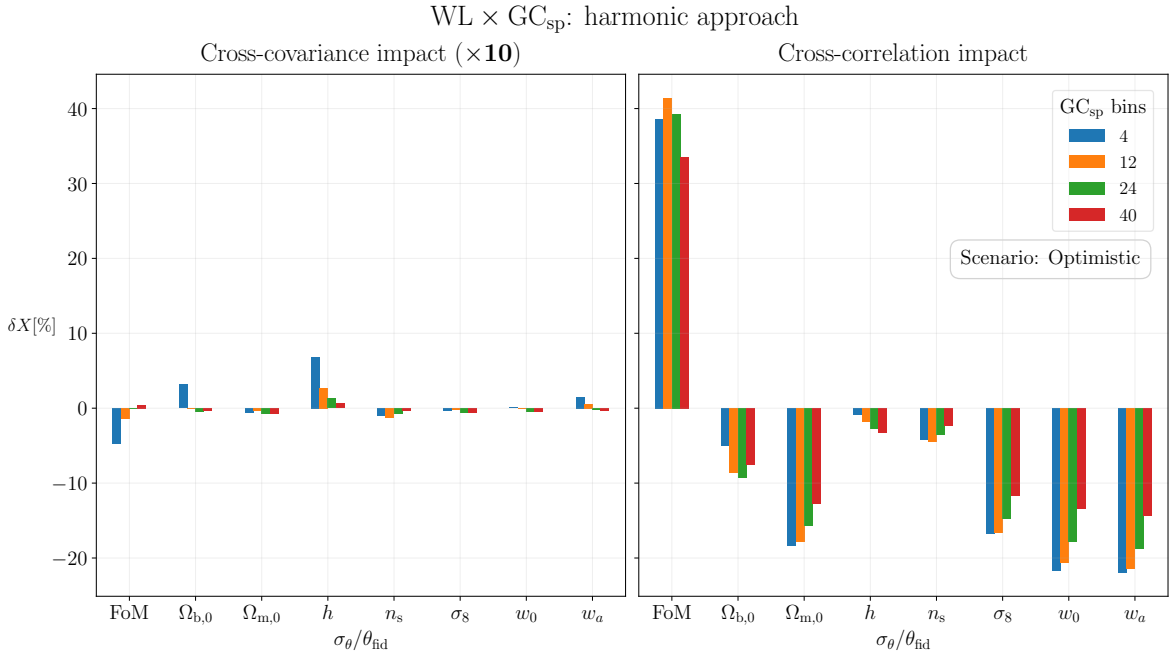

Figure 4.9: Impact, on 1–σ parameter errors and FoM, of cross-covariance and cross-correlation in the harmonic approach for the combination of WL and GC_{sp}. The left panel shows the impact of the cross-covariance, quantified with the percentage differences between [WL+GC_{sp}] and [WL]+[GC_{sp}]. In the right panel the impact of the cross-correlation is reported, quantified by percentage differences of [WL + GC_{sp} + XC(WL, GC_{sp})] with respect to [WL] + [GC_{sp}]. Note that the percentages on the left panel have been multiplied by 10 in order to make them visible with the same vertical scale of the right panel.

The same conclusion can be drawn by looking at the effect of the covariance on the marginalised 1–σ uncertainties on the cosmological parameters, shown in the left panel of Fig. 4.9. The marginalised uncertainties affected the most by the cross-covariance are the ones on h and $\Omega_{b,0}$, with variations of 0.7% and 0.3% respectively. The variations on all the other parameters are always well below 0.2%, and in all cases the absolute value of these variations decreases as the number of GC_{sp} bins increases, confirming the same trend observed for the FoM.

The conclusion here is that, when combining WL and GC_{sp} in the harmonic domain, their cross-covariance can be safely neglected. A similar result for another experiment other than *Euclid* has been obtained in [85]. Here the authors performed a joint data analysis combining weak lensing measurements from the Kilo-Degree Survey (KiDS-1000) and spectroscopic clustering from the Baryon Acoustic Oscillations Survey (BOSS) and 2-degree Field Lensing Survey (2dFLenS). The WL was treated using the harmonic power spectrum as observable, as it has been done in this work. Moreover, the cross-covariance matrix was computed in the harmonic domain, considering only the correlation between WL and the transverse component of GC_{sp}. The authors estimated the covariance matrix for the data through an analysis of over 20 000 fast full-sky mock galaxy catalogues, finding that the off-diagonal (cross-covariance) terms were negligible with respect to the diagonal (auto-covariance) ones.

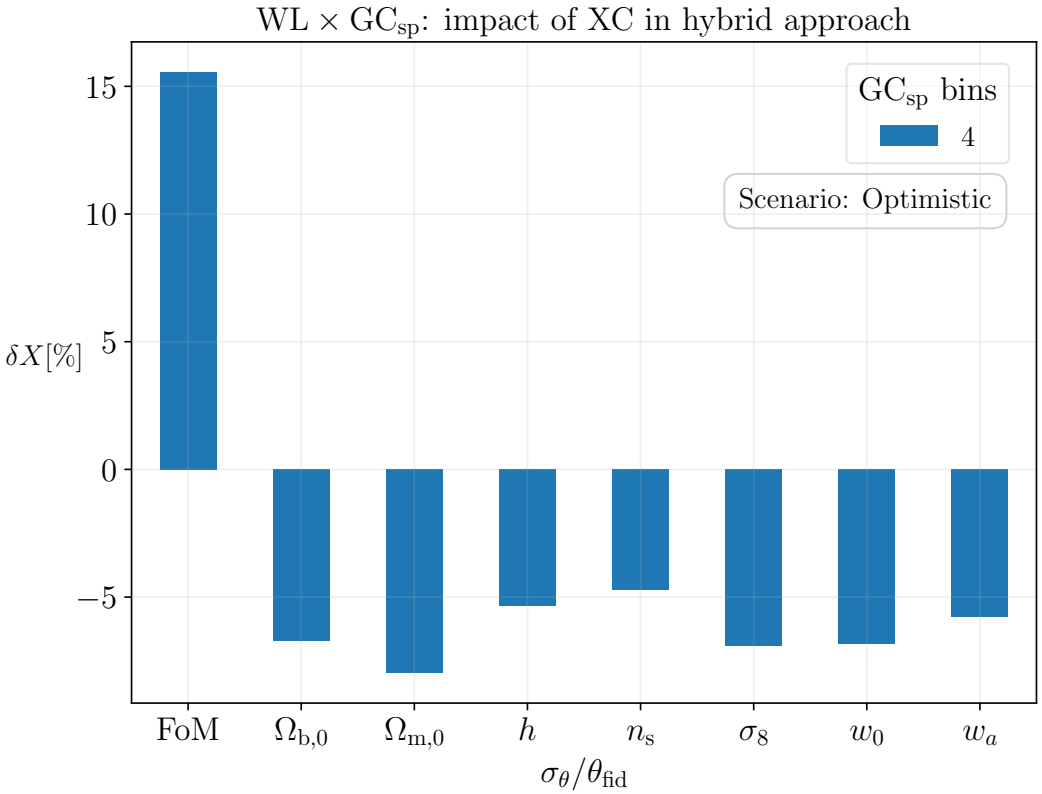

Figure 4.10: Impact, on FoM and marginalised 1–σ errors, of the XC(WL, GC$_{sp}$) in the hybrid approach, for the optimistic scenario. Note that in the hybrid approach only 4 tomographic bins were used to compute the XC(WL, GC$_{sp}$) angular power spectra, for consistency with the Fourier power spectrum analysis.

Impact of the XC(WL, GC$_{sp}$) signal on parameter constraints

Harmonic approach When combining WL and GC$_{sp}$ only, the XC(WL, GC$_{sp}$) has quite a significant impact on the results of the harmonic approach, as Tab. 4.7 shows. The percentage gain on the figure of merit is always larger than +30%, and slightly depends on the number of spectroscopic bins used. This can be explained by observing that WL weight functions have a broad support, which becomes larger as the tomographic index increases, as Fig. 4.1 shows. This suggests that increasing the radial resolution may not help in improving the constraints coming from the XC(WL, GC$_{sp}$). It is also worth noting that the FoM percentage gain does not strictly increase with the number of spectroscopic bins. In particular it is +38% at 4 bins and +33% with 40 bins, with a maximum of about +41% at 12 bins. This not intuitive behaviour is due to the normalisation of the FoM percentage difference, which is the FoM of the [WL] + [GC$_{sp}$] Fisher matrix. This quantity depends on the number of spectroscopic bins, and it increases slightly faster than the variation induced by the cross-correlation in the FoM of the [WL + GC$_{sp}$ + XC(WL, GC$_{sp}$)] Fisher matrix. This can be seen from Tab. 4.7: the difference ΔFoM between the FoMs of the [WL + GC$_{sp}$ + XC(WL, GC$_{sp}$)] and [WL] + [GC$_{sp}$] Fisher matrices grows slower with the number of bins than the FoM of [WL] + [GC$_{sp}$] alone.

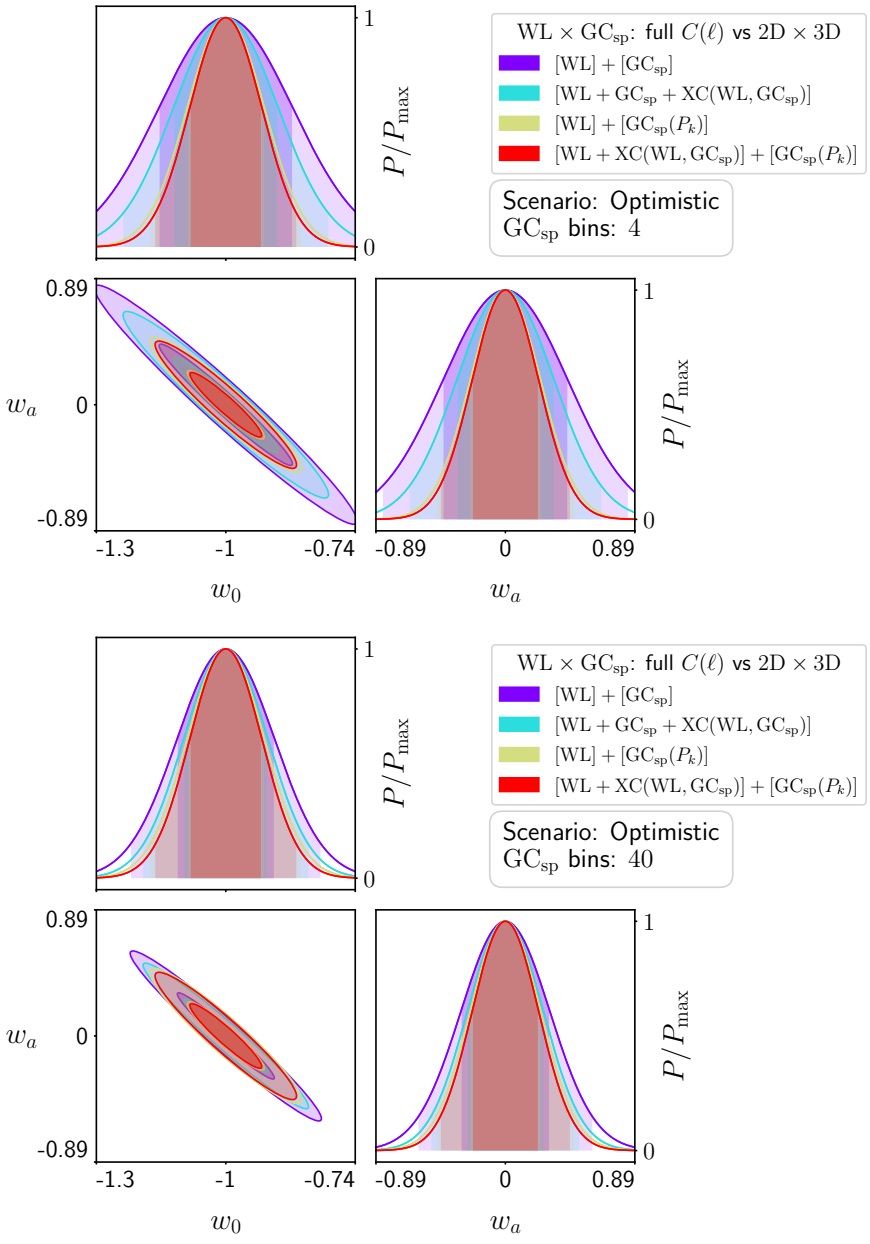

Figure 4.11: Contour plots showing the comparison between the two approaches in the combination of WL and GC$_{sp}$.

The marginalised 1–σ uncertainties decrease when the XC(WL, $\mathrm{GC_{sp}}$) cross-correlation signal is included, especially the ones on w_0, w_a, σ_8 and $\Omega_{\mathrm{m},0}$, as it can be seen from Fig. 4.9. For these parameters the improvement is always larger than 12% with 4 $\mathrm{GC_{sp}}$ bins, and it is about 20% at most when 40 bins are used. The gain on the $\Omega_{\mathrm{b},0}$ uncertainty is more modest, being comprised between 5% and 10%. The uncertainties on the reduced Hubble constant h and the scalar spectral index n_{s} gain always less than 5%.

The improvement on the marginalised uncertainties XC(WL, $\mathrm{GC_{sp}}$) cross-correlation seems to decrease when increasing the number of spectroscopic bins, as it happens for the FoM. Again, the reason of this behaviour is that the performances of the Fisher matrix taken as reference – that is [WL] + [$\mathrm{GC_{sp}}$] – improve faster than the relative contribution of the cross-correlation. This is true for all parameters except for h, which is the parameter affected the least.

Hybrid approach The impact on the constraints of the XC(WL, $\mathrm{GC_{sp}}$) cross-correlation in the hybrid approach is less significant than in the harmonic one, as shown in Fig. 4.10. The FoM of the [WL + XC(WL, $\mathrm{GC_{sp}}$)] + [$\mathrm{GC_{sp}}(P_k)$] Fisher matrix is 183, which is $\sim 15\%$ higher than the one of the independent combination [WL] + [$\mathrm{GC_{sp}}(P_k)$], which is 158. The improvements on the marginalised uncertainties are all comprised between 5% and 7%, and there are no significant differences between the various parameters. The uncertainty on $\Omega_{\mathrm{m},0}$ is the most affected, gaining about 8%, while the least affected is the uncertainty on n_{s}, which is slightly lower than 5%.

Hybrid approach vs harmonic approach

In contrast to what happens in the case of $\mathrm{GC_{ph}}$ and $\mathrm{GC_{sp}}$, when combining WL and $\mathrm{GC_{sp}}$ the harmonic approach can reach the FoM of the hybrid one. However, this only happens when 40 $\mathrm{GC_{sp}}$ bins are used to compute the harmonic Fisher matrix [WL + $\mathrm{GC_{sp}}$ + XC(WL, $\mathrm{GC_{sp}}$)]. The value reached by the FoM is 188, which is only 3% higher than the one given by the hybrid approach. Moreover, Fig. 4.12 shows that the marginalised uncertainties in w_0, w_a of the hybrid approach are smaller than their harmonic counterparts. Nonetheless, as said above, the FoM of the harmonic approach is slightly higher than the one of the hybrid approach. This is due to the correlation $\mathcal{C}_{w_0 w_a}$ between the parameters, which enters the definition [Eq. (4.27)] of the FoM. This correlation is higher for the harmonic approach, and this compensates for the larger uncertainties, with a net result of slightly higher FoM. This effect can also be seen by zooming on the bottom panel of Fig. 4.11. The contour of the harmonic approach (cyan) in the w_0-w_a plane is slightly narrower than the contour of the hybrid one (red), even if being more elongated, and this results in a smaller area of the contour ellipse – which means an higher FoM – of the former with respect to the latter.

Regarding the marginalised uncertainties, Fig. 4.12 shows that the hybrid approach always performs better than the harmonic one. The reduced Hubble constant h is the parameter for which the difference is the highest. In particular, the hybrid approach produces an uncertainty on h which is $\sim 90\%$ (75%) smaller than the one given by the harmonic approach with 4 (40) spectroscopic bins. For the uncertainties on the other parameters the gap is smaller, and it reduces significantly as the number of spectroscopic bins increases. The most sensitive uncertainty is the one on $\Omega_{\mathrm{b},0}$, for which the gap between the two approaches decreases from more than 50% to less than 10% when 4 and 40 bins are used for the harmonic approach respectively. The uncertainty on n_{s} is quite sensitive to the number of bins too, and the

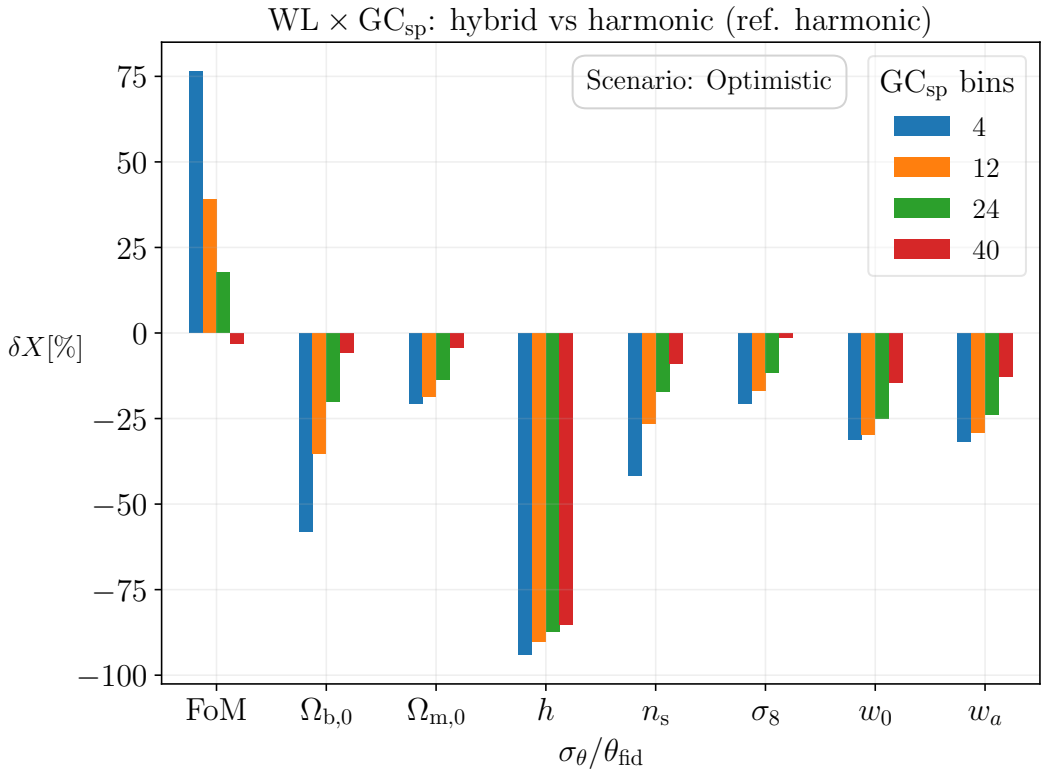

Figure 4.12: Percentage difference between FoM and marginalised 1–σ errors of the hybrid and harmonic approaches, for the combination of WL and GC$_{sp}$ in the baseline optimistic scenario. The percentage differences are normalised to the results of the harmonic approach.

difference between the two approaches ranges from 40% to 10% when the number bins of the harmonic approach increases from 4 to 40. Finally, the differences on the marginalised uncertainties on $\Omega_{m,0}$ and σ_8 are smaller, ranging from 20% to less than 5%.

Comparing XC(WL, GC$_{sp}$) with XC(WL, GC$_{ph}$)

In [5] it has been shown that the XC(WL, GC$_{ph}$) cross-correlation considerably improves the constraints on the cosmological parameters. In this work the same result is found: in the optimistic scenario the Fisher matrix [WL + GC$_{ph}$ + XC(WL, GC$_{ph}$)] yields a FoM which is a factor of ~ 5 higher than the one given by the [WL] + [GC$_{ph}$] Fisher matrix. This means that the percentage gain induced by the cross-correlation signal is about 400%. The XC(WL, GC$_{sp}$) cross-correlation has a smaller impact on the constraints, as the [WL+GC$_{sp}$+XC(WL, GC$_{sp}$)] Fisher matrix produces a FoM at most $\sim$ 40% higher than the one of the [WL] + [GC$_{sp}$] combination. This is what happens when both GC$_{ph}$ and GC$_{sp}$ are treated in the baseline settings, i.e. when the standard shot noise Eq. (4.49) is used for GC$_{sp}$ and 10 tomographic bins are used for GC$_{ph}$. The effect of the shot noise on GC$_{sp}$ can be seen from Fig. 4.13, where I report the percentage gain on the FoM of various combinations with respect to WL alone. Left panel refers to the standard GC$_{sp}$ shot noise computed with the second equation of Eq. (4.49), right panel refers to the alternative reduced noise of Eq. (4.53). When the noise

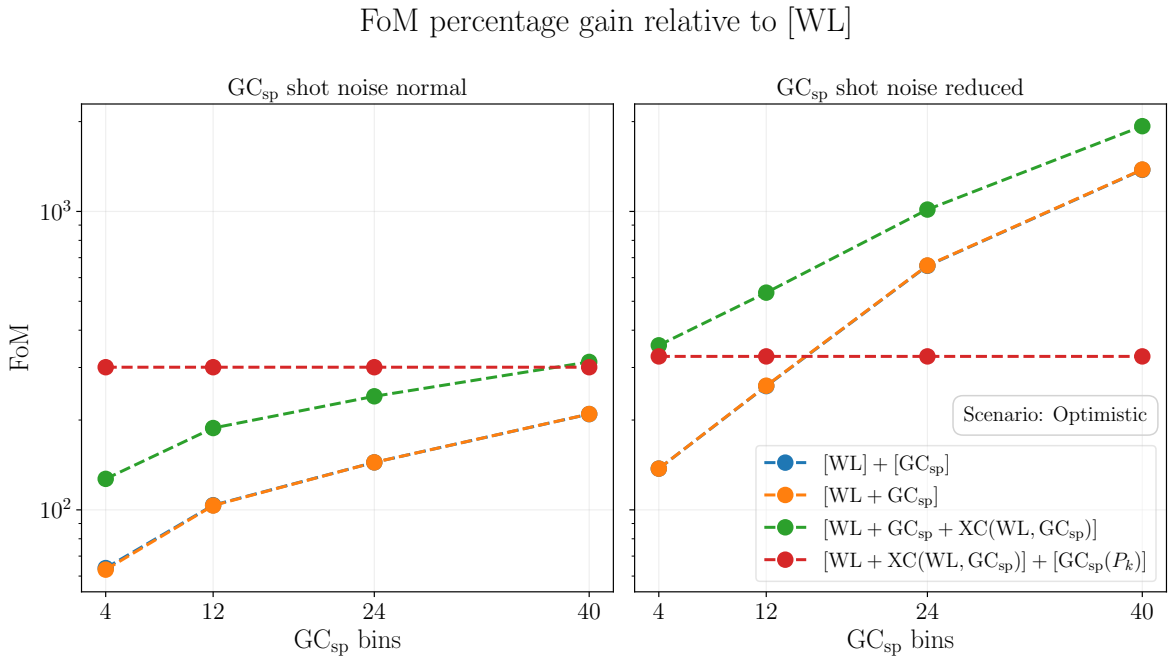

Figure 4.13: Percentage gain on the FoM relative to weak lensing clustering alone ([WL]). In the left panel the computation has been done using the standard $\mathrm{GC_{sp}}$ shot noise Eq. (4.49), while in the right panel the reduced version Eq. (4.53) has been used. Note that $[\mathrm{WL}]+[\mathrm{GC_{sp}}]$ (blue points) is indistinguishable from $[\mathrm{WL}+\mathrm{GC_{sp}}]$ (orange points), confirming that the cross-covariance between WL and $\mathrm{GC_{sp}}$ is negligible.

is reduced the Fisher matrix of the harmonic approach with 4 bins is already competitive with the one of the hybrid approach, yielding a gain of $\sim 300\%$ relative to WL alone. The observable gaining the most from the noise reduction is the $\mathrm{GC_{sp}}$ auto-correlation, as the two curves of $[\mathrm{WL}+\mathrm{GC_{sp}}]$ and $[\mathrm{WL}+\mathrm{GC_{sp}}+\mathrm{XC}(\mathrm{WL},\mathrm{GC_{sp}})]$ converges towards each other as the number of bins increases.

A direct comparison between the gain coming from $\mathrm{XC}(\mathrm{WL},\mathrm{GC_{ph}})$ and $\mathrm{XC}(\mathrm{WL},\mathrm{GC_{sp}})$ cross-correlations is shown in Fig. 4.14, where the percentage variations on parameter constraints are represented with vertical bars. When $\mathrm{GC_{sp}}$ and $\mathrm{GC_{ph}}$ have the same shot noise level and the same redshift range, the effect of the $\mathrm{XC}(\mathrm{WL},\mathrm{GC_{sp}})$ on the constraints becomes comparable with the one of $\mathrm{XC}(\mathrm{WL},\mathrm{GC_{ph}})$.

Therefore, the conclusion is the same drawn for the direct comparison between the $\mathrm{GC_{ph}}$ and $\mathrm{GC_{sp}}$ auto-correlations. The shot noise and the redshift range of the galaxy catalogue are what makes the differences between $\mathrm{XC}(\mathrm{WL},\mathrm{GC_{ph}})$ and $\mathrm{XC}(\mathrm{WL},\mathrm{GC_{sp}})$ in terms of constraining power.

4.3.4 Exploiting the *Euclid* 6×2pt statistics

In this section I present the results of the combination of all the *Euclid* main probes: $\mathrm{WL}, \mathrm{GC_{ph}}$, and $\mathrm{GC_{sp}}$. The starting point is the (photometric) 3×2pt statistics, defined

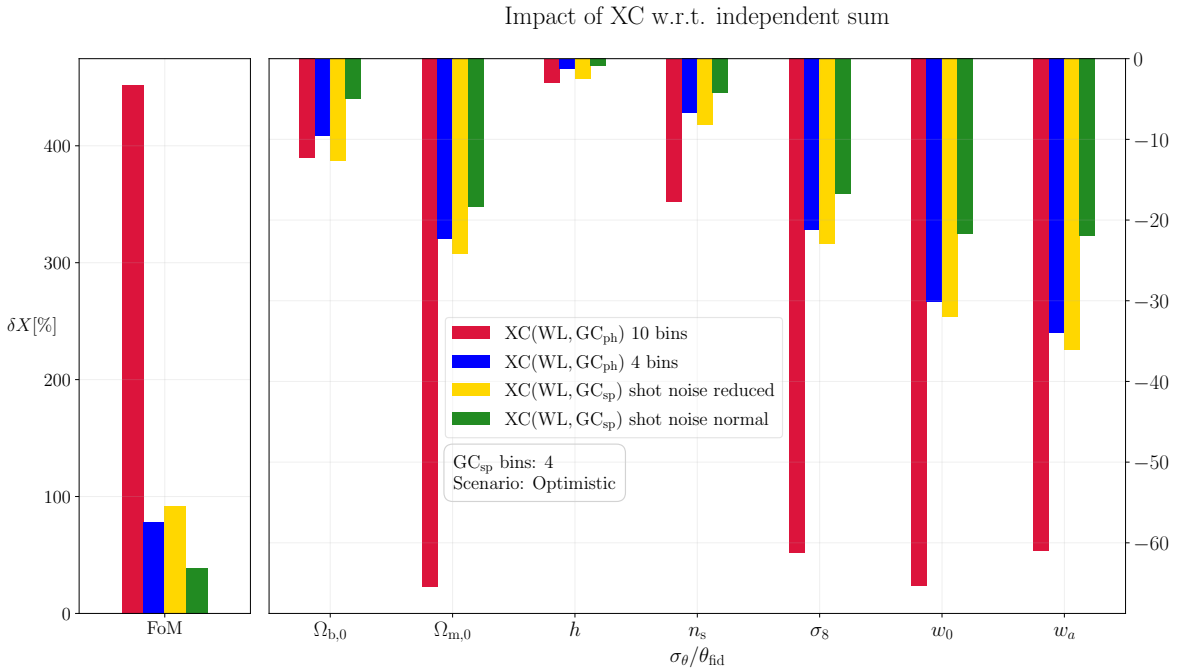

Figure 4.14: Impact, on FoM and marginalised 1–σ uncertainties, of the XC(WL, GC$_{ph}$) and XC(WL, GC$_{sp}$) cross-correlations in different configurations. The vertical bars represent the percentage differences on the uncertainties normalised to the independent sum of the two probes. For XC(WL, GC$_{ph}$) the reference is [WL]+[GC$_{ph}$], while for XC(WL, GC$_{sp}$) is [WL]+[GC$_{sp}$]. Note that for the uncertainties the percentage variation is *negative*, meaning that the contribution of the cross-correlations *reduces* the errors.

as

$$3\times 2\text{pt} \equiv [\text{WL} + \text{GC}_{\text{ph}} + \text{XC}(\text{WL}, \text{GC}_{\text{ph}})]\,. \tag{4.60}$$

This combination has already been studied in [5], finding that the XC(WL, GC$_{ph}$) cross-correlation has a large impact on the constraints. The same result is found in this work too, as it has been already pointed out in Sec. 4.3.3 – see the red bars of Fig. 4.14. In terms of this combination, the 6×2pt statistics, defined in Eq. (4.58) and Eq. (4.59), can be expressed as

$$\begin{aligned}[(\text{harmonic})\, 6\times 2\text{pt}] &= [3\times 2\text{pt} + \text{GC}_{\text{sp}} + \text{XC}(\text{WL}, \text{GC}_{\text{sp}}) + \text{XC}(\text{GC}_{\text{ph}}, \text{GC}_{\text{sp}})] \\ [(\text{hybrid})\, 6\times 2\text{pt}] &= [3\times 2\text{pt} + \text{XC}(\text{WL}, \text{GC}_{\text{sp}}) + \text{XC}(\text{GC}_{\text{ph}}, \text{GC}_{\text{sp}})] + [\text{GC}_{\text{sp}}(P_k)]\end{aligned} \tag{4.61}$$

In this section I report the *Euclid* performance when GC$_{sp}$ and its cross-correlations with WL, and GC$_{ph}$ are combined with the 3×2pt statistics. The discussion will be focused on two main points:

- the importance of the cross-covariance between GC$_{sp}$ and the 3×2pt statistics;
- the contribution to the constraints of the XC(GC$_{ph}$, GC$_{sp}$) and XC(WL, GC$_{sp}$) cross-correlations.

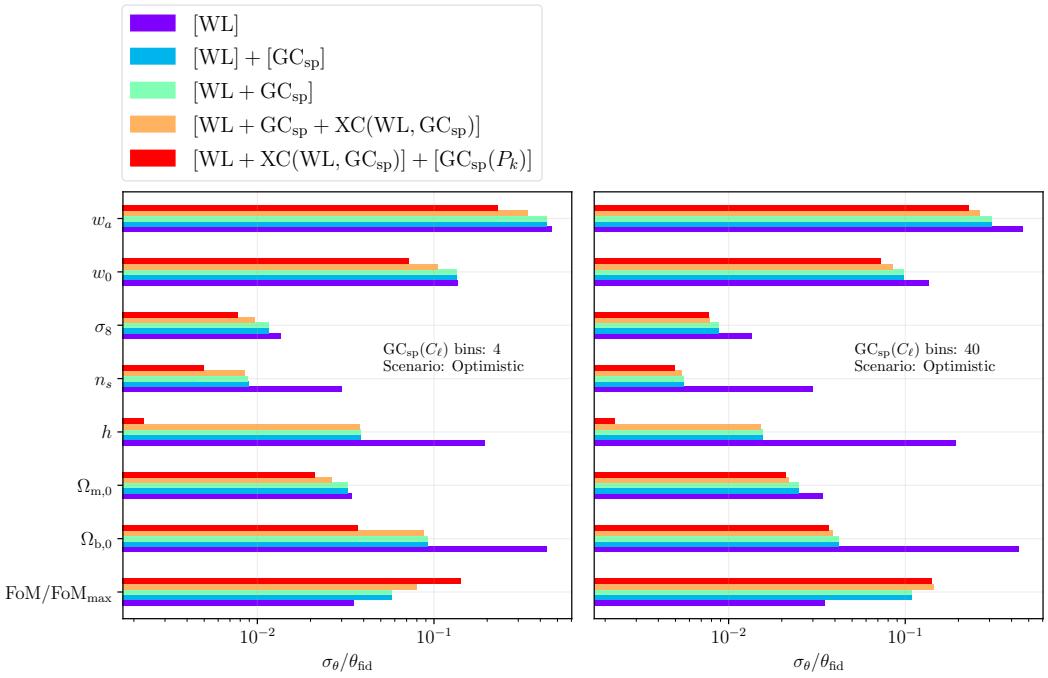

Figure 4.15: Comparison between harmonic and hybrid harmonic-Fourier approaches for the case of WL $\times$ GC$_{sp}$.

The cross-covariance between GC$_{sp}$ and 3×2pt statistics is studied only in the harmonic approach, since in the hybrid approach it is neglected by definition. The effect of the cross-correlations is assessed for both the harmonic and the hybrid approaches, and in both the pessimistic and optimistic scenarios defined in Tab. 4.3.1.

Impact of the GC$_{sp}$ cross-covariances on parameter constraints

The impact on the constraints of the cross-covariance between GC$_{sp}$ and 3×2pt is shown in the two panels of Fig. 4.16, the left one representing the optimistic scenario and the right the pessimistic one. The plot compares the constraints from the $[3 \times 2\text{pt} + \text{GC}_{sp}]$ and $[3 \times 2\text{pt}] + [\text{GC}_{sp}]$ combinations: in the former the GC$_{sp}$-3×2pt cross-covariance is taken into account, while it is not in the latter. It can be seen that the impact of the cross-covariance is almost the same in the two scenarios and decreases as the number of GC$_{sp}$ bins increases, confirming the same trend observed in the two pairwise combinations GC$_{ph}$-GC$_{sp}$ and WL-GC$_{sp}$. The percentage variations on the constraints are always below 10% (5%) with 4 (40) spectroscopic bins. It is also evident that the covariance almost always worsens the constraints, reducing the FoM and increasing the marginalised 1–σ uncertainties with respect to considering GC$_{sp}$ and 3×2pt as independent. The parameters whose uncertainties are affected the most by the GC$_{sp}$-3×2pt cross-covariance are $\Omega_{b,0}$ and h, with variations of $\sim 8\%$ with 4 bins. When 40 bins are used for GC$_{sp}$ the variation reduces to $\sim 2\%$ for both parameters.

On of the most evident differences between the optimistic and the pessimistic scenario is the impact on the FoM, which is slightly higher in the pessimistic than in the optimistic setting. However the percentage difference is always below 5%, the worst case being the pessimistic

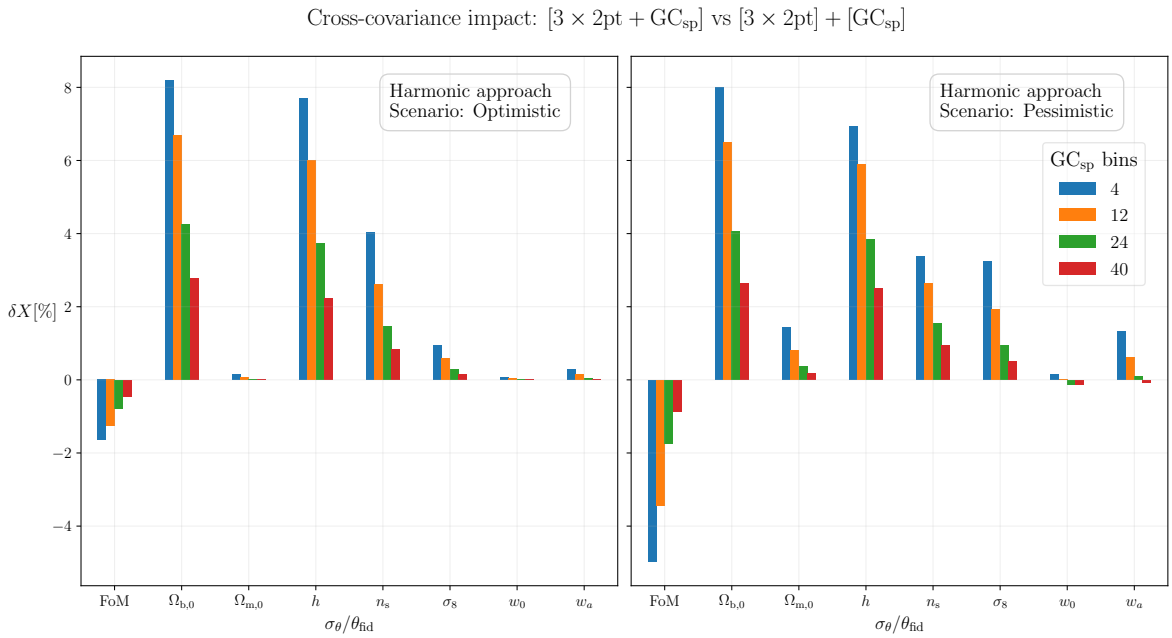

Figure 4.16: Impact, on FoM and marginalised 1-σ uncertainties, of cross covariances between GC_sp and $3\times2\text{pt}$ ($[\text{WL}+\text{GC}_\text{ph}+\text{XC}(\text{WL},\text{GC}_\text{ph})]$), quantified with percentage differences on the constraints, in the optimistic (left) and pessimistic (right) scenarios. The reference for the percentage are the constraints of $[3\times2\text{pt}]+[\text{GC}_\text{sp}]$, where GC_sp is considered as *independent* from the rest.

scenario with 4 bins, where it is $\sim 4\%$. The percentage variation on the $\Omega_{\text{m},0}$ uncertainty is at the sub-percent level in the optimistic scenario, while it is about at the percent level in the pessimistic scenario. The uncertainty on σ_8 is always smaller than 1% in the optimistic scenario, while it ranges from 4% to 1% in the pessimistic scenario.

Impact of the XC signals on parameter constraints

Here I discuss the impact on the constraints of the $\text{XC}(\text{GC}_\text{ph},\text{GC}_\text{sp})$ and $\text{XC}(\text{WL},\text{GC}_\text{sp})$ cross-correlations. The impact of the cross-correlations is quantified as percentage differences on the constraints with respect to the given reference, represented with vertical bars. Fig. 4.17 and Fig. 4.18 show the results in the harmonic approach, and Fig. 4.19 in the hybrid one. Each of these plots contains three columns. The left column reports the effect of the inclusion of the $\text{XC}(\text{WL},\text{GC}_\text{sp})$ cross-correlation, the middle column reports the impact of the $\text{XC}(\text{GC}_\text{ph},\text{GC}_\text{sp})$ cross-correlation, and the right column reports the impact of both.

Harmonic approach For the harmonic approach, two Fisher matrices have been used as reference for the percentage differences: $[3\times2\text{pt}]+[\text{GC}_\text{sp}]$, where the $3\times2\text{pt}$ statistics and GC_sp are combined as independent, and $[3\times2\text{pt}+\text{GC}_\text{sp}]$, where the cross-covariance between the two is taken into account (see Sec. 4.2.1). As explained in the previous section, these two combinations do not produce the same constraints (see Fig. 4.16), as the independent combination yields slightly better constraints.

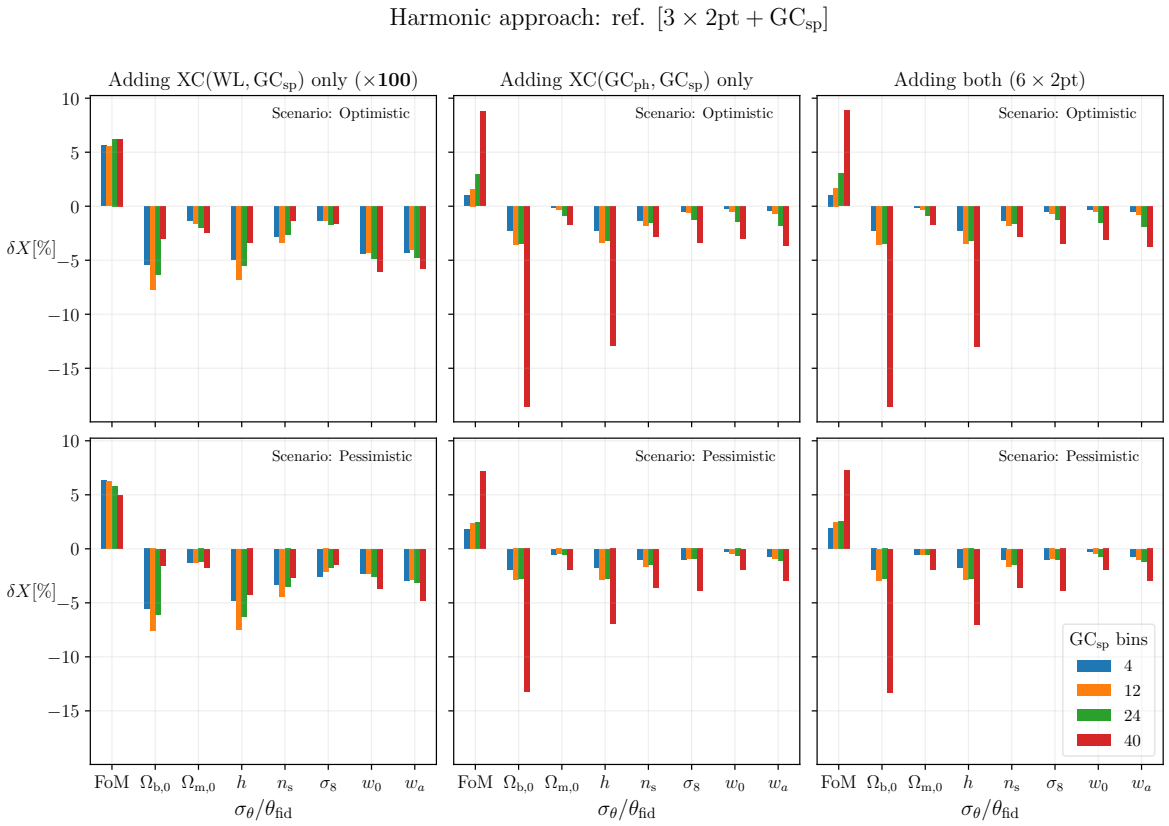

Figure 4.17: Impact of the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ and $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$ cross-correlations on FoM and marginalised 1–σ uncertainties, with respect to the constraints given by the $[3 \times 2\mathrm{pt} + \mathrm{GC_{sp}}]$ Fisher matrix, in which the cross-covariance between $3 \times 2\mathrm{pt}$ statistics and $\mathrm{GC_{sp}}$ is taken into account. The top panels refer to the optimistic scenario, while the bottom panels refer to the pessimistic one. Note that the percentage differences related to $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$ have been multiplied by 100 to make them visible when using a single scale on the y axis.

I consider the two different references above since, when the cross-correlations are added in the harmonic approach, the cross-covariance between $\mathrm{GC_{sp}}$ and the $3 \times 2\mathrm{pt}$ statistics is always accounted for. Thus, on the one hand, I focus on the improvement due to the inclusion of the XC signals alone, and this is done when the Fisher matrix $[3 \times 2\mathrm{pt} + \mathrm{GC_{sp}} + \mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})]$ is compared to the reference $[3 \times 2\mathrm{pt} + \mathrm{GC_{sp}}]$, and the same for $[3 \times 2\mathrm{pt} + \mathrm{GC_{sp}} + \mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})]$ and $[3 \times 2\mathrm{pt} + \mathrm{GC_{sp}} + \mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}}) + \mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})]$ (see Fig. 4.17). In this case, the percentage differences are representative of the *net effect* of the cross-correlation signals on the constraints, which is expected to be always positive.

On the other hand, when the Fisher matrices comprising the XC's information are compared to the independent combination $[3 \times 2\mathrm{pt}] + [\mathrm{GC_{sp}}]$, I focus on the *total effect*, which not only contains the gain from the inclusion of the XC signals, but also the penalty from the cross-covariance between $\mathrm{GC_{sp}}$ and the $3 \times 2\mathrm{pt}$ statistics (see Fig. 4.18). This comparison is useful in order to evaluate the overall impact of the cross angular power spectra.

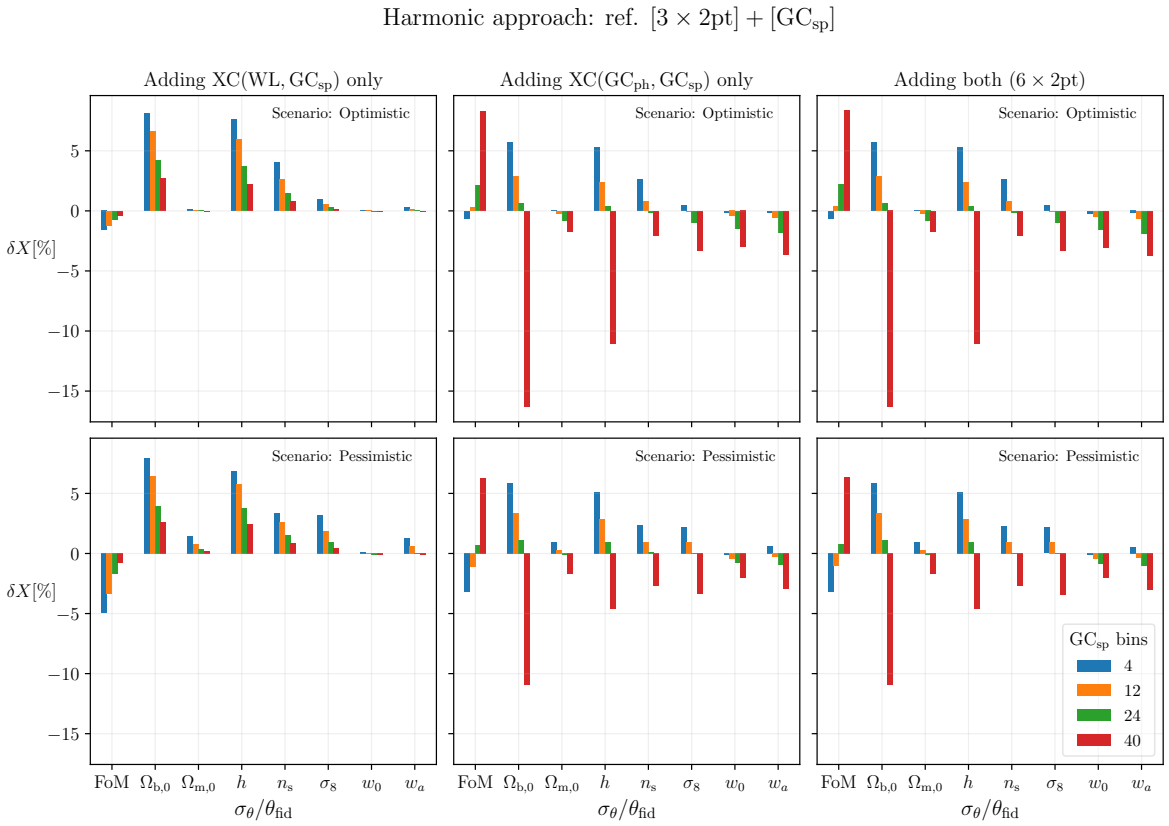

Figure 4.18: Impact of the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ and $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$ cross-correlations on FoM and marginalised 1–σ uncertainties, with respect to the constraints given by the $[3 \times 2\mathrm{pt}] + [\mathrm{GC_{sp}}]$ combination, where $3 \times 2\mathrm{pt}$ and $\mathrm{GC_{sp}}$ are treated as independent. The top panels refer to the optimistic scenario, while the bottom panels refer to the pessimistic one.

Harmonic approach: adding $\mathbf{XC(WL, GC_{sp})}$ The net effect of the $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$ inclusion is shown in the left panels of Fig. 4.17, where the Fisher matrix $[3 \times 2\mathrm{pt} + \mathrm{GC_{sp}} + \mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})]$ is compared to the reference $[3 \times 2\mathrm{pt} + \mathrm{GC_{sp}}]$. The differences in this case have been amplified by a factor of 100 to make them visible on the same vertical scale of the other plots. The variation on the constraints due to the addition of $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$ to $[3 \times 2\mathrm{pt} + \mathrm{GC_{sp}}]$ is always about 0.01%-0.05%, with no significant differences between the optimistic and pessimistic scenarios. Therefore, it seems that this cross-correlation does not give any contribution to the constraints, i.e. it looks like the computation of the total $[3 \times 2\mathrm{pt} + \mathrm{GC_{sp}} + \mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})]$ Fisher matrix is not useful to improve the *Euclid* performance. This might seem in contrast to what happens in the pairwise combination of WL and $\mathrm{GC_{sp}}$, where the $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$ cross-correlation signal has a significant impact on the constraints. However, the latter case does not include by definition the $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{ph}})$ signal, since it refers only to WL combined with $\mathrm{GC_{sp}}$. Instead, in the case now under discussion of $[3 \times 2\mathrm{pt} + \mathrm{GC_{sp}} + \mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})]$, the reference for the net effect here considered is $[3 \times 2\mathrm{pt} + \mathrm{GC_{sp}}]$, which contains the contribution of the $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{ph}})$ cross-correlation, proven to be dominant.

From the above reasoning, it is also possible to infer the reason why the constraints from

the $[3\times2\text{pt}+\text{GC}_\text{sp}+\text{XC}(\text{WL},\text{GC}_\text{sp})]$ Fisher matrix are worse by 6% at most (see Fig. 4.18) than the ones from the independent combination $[3\times2\text{pt}]+[\text{GC}_\text{sp}]$ taken now as reference: the impact of the cross-correlation signal is so small that its possible improvements are completely dominated by the cross-covariance between GC_sp and the $3\times2\text{pt}$ statistics (see Fig. 4.16), which is present when adding the cross signal, but not kept into account in the reference. Therefore, the total effect of the inclusion of $\text{XC}(\text{WL},\text{GC}_\text{sp})$ in the combination of GC_sp with the $3\times2\text{pt}$ statistic is to worsen the parameter constraints. However, while it has been just shown that the $\text{XC}(\text{WL},\text{GC}_\text{sp})$ signal can be safely neglected, the cross-covariance between GC_sp and the $3\times2\text{pt}$ statistic needs to be taken with caution.

Harmonic approach: adding $\mathbf{XC(GC_{ph},GC_{sp})}$ The middle panel of Fig. 4.17 shows the *positive* net effect of the $\text{XC}(\text{GC}_\text{ph},\text{GC}_\text{sp})$ inclusion, which increases with the number of spectroscopic bins. This confirms the same behaviour observed in the pairwise combination of GC_ph and GC_sp. The gain on the FoM relative to the Fisher matrix $[3\times2\text{pt}+\text{GC}_\text{sp}]$ is about 1% for 4 bins, and increases up to 8-10% for 40 bins, with practically no differences between the optimistic and the pessimistic scenario. Fig. 4.18 shows that when the independent combination $[3\times2\text{pt}]+[\text{GC}_\text{sp}]$ is used as reference instead, the FoM variation due to the $\text{XC}(\text{GC}_\text{ph},\text{GC}_\text{sp})$ inclusion is -2% ($+7\%$) for 4 (40) GC_sp bins. The small worsening at 4 bins is due to the fact that the positive contribution of the cross-correlation is cancelled by the negative contribution of the GC_sp-$3\times2\text{pt}$ cross-covariance (see Fig. 4.16). In fact, this cross-covariance is taken into account in the Fisher matrix $[3\times2\text{pt}+\text{GC}_\text{sp}+\text{XC}(\text{GC}_\text{ph},\text{GC}_\text{sp})]$, while it is not in the $[3\times2\text{pt}]+[\text{GC}_\text{sp}]$, which is used as the reference in this last case.

Analogously, the marginalised 1–σ uncertainties on the cosmological parameters exhibit a similar behaviour, with no significant differences between the optimistic and the pessimistic scenarios. When using 4 spectroscopic bins the inclusion of the $\text{XC}(\text{GC}_\text{ph},\text{GC}_\text{sp})$ cross-correlation produces a small improvement when the Fisher matrix $[3\times2\text{pt}+\text{GC}_\text{sp}]$ is used as reference. For a small number of bins, this positive contribution is in general compensated by cross-covariance effects when the percentage differences are referred to the $[3\times2\text{pt}]+[\text{GC}_\text{sp}]$ Fisher matrix. When using 40 spectroscopic bins the cross-correlation dominates and the cross-covariance effects become negligible, and the percentage differences become always positive, independently of the reference that is used. The parameters whose uncertainties decrease the most are $\Omega_{\text{b},0}$ and h, gaining 15% and 10% respectively in the optimistic scenario, 10% and 5% in the pessimistic.

Therefore, the total effect of the inclusion of $\text{XC}(\text{GC}_\text{ph},\text{GC}_\text{sp})$ in the combination of GC_sp with the $3\times2\text{pt}$ statistic depends on the chosen binning set, and may be dominant with respect to GC_sp-$3\times2\text{pt}$ cross-covariance effects for a large number of bins.

Harmonic approach: the $\mathbf{6\times2pt}$ statistics The harmonic $6\times2\text{pt}$ statistics in Eq. (4.58) consists in the inclusion of the $\text{XC}(\text{WL},\text{GC}_\text{sp})$ and $\text{XC}(\text{GC}_\text{ph},\text{GC}_\text{sp})$ cross-correlations in the $[3\times2\text{pt}+\text{GC}_\text{sp}]$ data-vector. In the harmonic approach, the constraints produced by the $6\times2\text{pt}$ analysis are equivalent to the ones given by including the $\text{XC}(\text{GC}_\text{ph},\text{GC}_\text{sp})$ alone. The percentage differences between the constraints from the $6\times2\text{pt}$ Fisher matrix and the $[3\times2\text{pt}+\text{GC}_\text{sp}]$ Fisher matrix are reported in the rightmost panels of Fig. 4.17. These are indistinguishable from the ones reported in the middle panels, which refer to the impact of the $\text{XC}(\text{GC}_\text{ph},\text{GC}_\text{sp})$ cross-correlation only with respect to $[3\times2\text{pt}+\text{GC}_\text{sp}]$. This is expected,

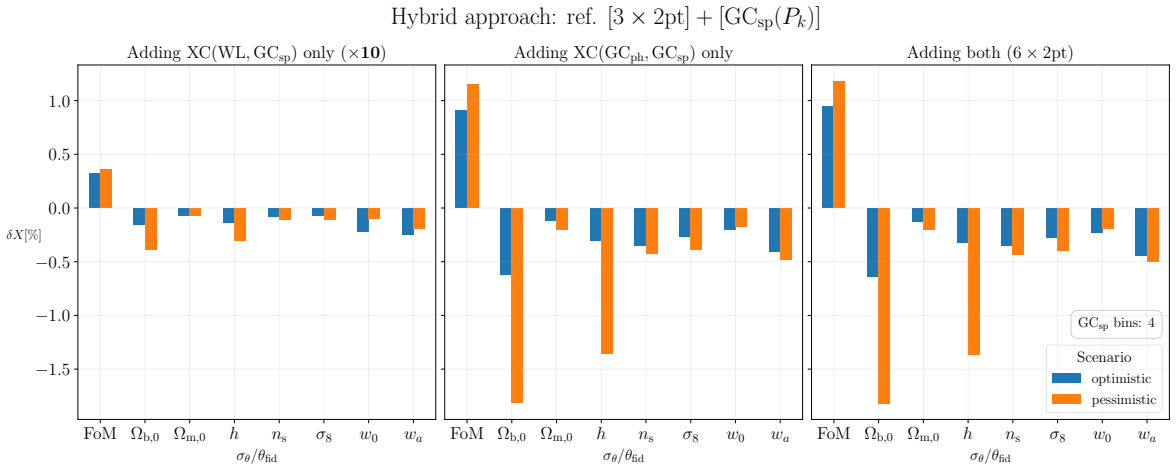

Figure 4.19: Impact, on FoM and marginalised 1–σ errors, of the XC($\text{GC}_{\text{ph}}, \text{GC}_{\text{sp}}$) and XC($\text{WL}, \text{GC}_{\text{sp}}$) cross-correlations in the hybrid approach. The reference for the percentage differences is $[\text{WL} + \text{GC}_{\text{ph}} + \text{XC}(\text{WL}, \text{GC}_{\text{ph}})] + [\text{GC}_{\text{sp}}(P_k)]$. In any case the cross-correlations are always considered to be *covariant* with the 3×2pt statistics observables. Note that the percentage differences in left panel have been multiplied by 10, to make them visible to the naked eye.

since in the above paragraphs it has been shown that XC($\text{WL}, \text{GC}_{\text{sp}}$) provides a negligible contribution with respect to XC($\text{GC}_{\text{ph}}, \text{GC}_{\text{sp}}$).

Hybrid approach The contribution of the XC($\text{WL}, \text{GC}_{\text{sp}}$) and XC($\text{GC}_{\text{ph}}, \text{GC}_{\text{sp}}$) signals in the hybrid approach is reported in Fig. 4.19, quantified with percentage differences of the constraints plotted as vertical bars. In this case the reference used is always the independent combination $[3 \times 2\text{pt}] + [\text{GC}_{\text{sp}}(P_k)]$, since in the hybrid approach the cross-covariance between $\text{GC}_{\text{sp}}(P_k)$ and the 3×2pt is neglected by definition. Only the 4 bins case is presented for consistency with the power spectrum analysis done in [5], for which the authors employ the 4 standard spectroscopic bins listed in Tab. 4.4. Again, Fig. 4.19 is subdivided in three columns. The first column reports the impact of XC($\text{WL}, \text{GC}_{\text{sp}}$), the middle column reports the impact of XC($\text{GC}_{\text{ph}}, \text{GC}_{\text{sp}}$), and finally in the last column reports the impact of both.

Hybrid approach: adding XC($\text{WL}, \text{GC}_{\text{sp}}$) In the hybrid approach the impact of the XC($\text{WL}, \text{GC}_{\text{sp}}$) cross-correlation on the constraints from $[3 \times 2\text{pt} + \text{XC}(\text{WL}, \text{GC}_{\text{sp}})] + [\text{GC}_{\text{sp}}(P_k)]$, relative to $[3 \times 2\text{pt}] + [\text{GC}_{\text{sp}}(P_k)]$, is negligible, being always less than 0.05%, both in the optimistic and in the pessimistic scenarios. The contribution is so small that it has been amplified by a factor of 10, in order to make it visible when using a single scale on the vertical axis. This result is analogous to what has been found with the harmonic approach, when I consider the comparison of $[3 \times 2\text{pt} + \text{GC}_{\text{sp}} + \text{XC}(\text{WL}, \text{GC}_{\text{sp}})]$ with respect to $[3 \times 2\text{pt} + \text{GC}_{\text{sp}}]$, to isolate the impact of XC($\text{WL}, \text{GC}_{\text{sp}}$). This is expected, since there are no differences in the 3×2pt between the two approaches, and, in particular, the XC($\text{WL}, \text{GC}_{\text{ph}}$) cross-correlation is computed in the same way in the two cases.

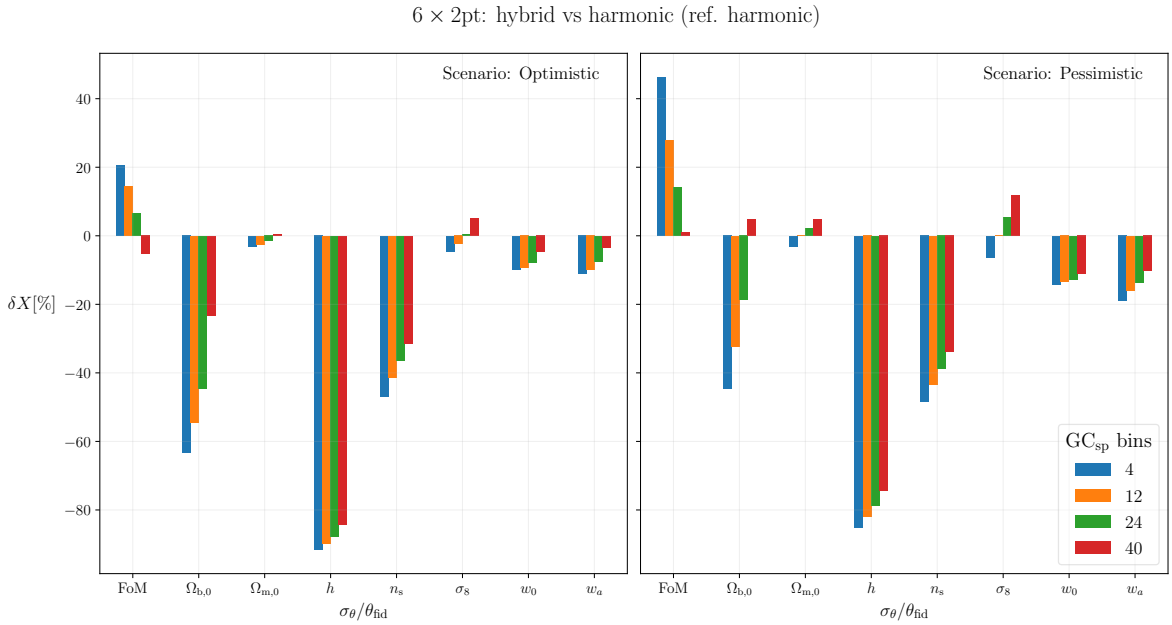

Figure 4.20: Comparison between FoM and marginalised 1–σ uncertainties of the hybrid 6×2pt and the harmonic 6×2pt statistics, quantified using percentage differences taking the latter as reference.

Hybrid approach: adding $\mathbf{XC(GC_{ph}, GC_{sp})}$ In the hybrid approach the impact of the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ cross-correlation on the constraints from $[3\times2\mathrm{pt} + \mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})] + [\mathrm{GC_{sp}}(P_k)]$, relative to $[3\times2\mathrm{pt}] + [\mathrm{GC_{sp}}(P_k)]$, is slightly smaller than in the harmonic one with 4 spectroscopic bins. The middle panel of Fig. 4.19 shows that the absolute percentage differences on all constraints is always below 2%. The gain on the FoM is $\sim 1\%$ both in the optimistic and the pessimistic scenario. The parameters whose uncertainties are affected the most are $\Omega_{\mathrm{b},0}$ and h, with a gain of 1.5% at most in the pessimistic scenario, and less than 0.5% in the optimistic scenario.

Hybrid approach: the $\mathbf{6\times2}$pt statistics As it happens for the harmonic approach, the constraints given by the hybrid 6×2pt statistics are similar to the ones given by the inclusion of the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ cross-correlation only. This is manifest in the right panel of Fig. 4.19: the vertical bars representing the differences are indistinguishable from the ones in the middle panel. Nonetheless, in the hybrid approach the inclusion of the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ cross-correlation has a negligible impact on the constraints, as discussed in the above paragraph. Therefore, the Fisher matrix of the hybrid 6×2pt statistics produces constraints that are almost equivalent to the $[3\times2\mathrm{pt}] + [\mathrm{GC_{sp}}(P_k)]$ Fisher matrix.

The 6×2pt statistics: hybrid approach vs harmonic approach

For the 6×2pt statistics the hybrid approach performs better than the harmonic one, especially when a small number of spectroscopic bins is used for the latter. The comparison of the two approaches is reported in Fig. 4.20, where the percentage differences between their constraints

6×2pt FoM results				
GC_{sp} bins	Fisher matrix	FoM	ΔFoM	ΔFoM(%)
	Optimistic scenario			
4	$[3 \times 2pt] + [GC_{sp}(P_k)]$	1216.16	–	–
	$[6 \times 2pt$ (hybrid)]	1227.69	11.53	+0.95%
	$[6 \times 2pt$ (harmonic)]	1018.43	−197.73	−16.26%
12	$[6 \times 2pt$ (harmonic)]	1073.31	−142.85	−11.75%
24	$[6 \times 2pt$ (harmonic)]	1151.13	−65.03	−5.35%
40	$[6 \times 2pt$ (harmonic)]	1296.44	80.28	+6.60%
	Pessimistic scenario			
4	$[3 \times 2pt] + [GC_{sp}(P_k)]$	549.37	–	–
	$[6 \times 2pt$ (hybrid)]	555.87	6.50	+1.18%
	$[6 \times 2pt$ (harmonic)]	379.85	−169.52	−30.86%
12	$[6 \times 2pt$ (harmonic)]	434.00	−115.37	−21.00%
24	$[6 \times 2pt$ (harmonic)]	486.40	−62.97	−11.46%
40	$[6 \times 2pt$ (harmonic)]	550.82	1.45	+0.26%

Table 4.8: Figure-of-Merit of the 6×2pt statistics in the harmonic and hybrid approaches and in the hybrid one. The ΔFoM columns quantify the differences with respect to the $[3 \times 2pt] + [GC_{sp}(P_k)]$ combination, which is used as reference to assess the impact on the FoM of the $XC(GC_{ph}, GC_{sp})$ and $XC(WL, GC_{sp})$ cross-correlations. In fact this combination does not include the contributions of $XC(GC_{ph}, GC_{sp})$ and $XC(WL, GC_{sp})$, but only the one of $XC(WL, GC_{ph})$ – see Eq. (4.60). For the definitions of 6×2pt statistics in the two approaches see Eq. (4.58) and Eq. (4.59).

are plotted as vertical bars. In this case, the reference adopted is the Fisher matrix of the harmonic approach, Eq. (4.58). The hybrid approach produces a FoM which is 20% (40%) larger than the harmonic one in the optimistic (pessimistic) scenario, when 4 spectroscopic bins are used for the latter. When using 40 bins, the harmonic approach performs instead slightly better ($\sim 6\%$) than the hybrid one in the optimistic scenario, while it is equivalent to it in the pessimistic scenario.

Regarding the marginalised 1–σ uncertainties on the dark energy parameters, w_0, w_a, the hybrid approach always provides better constraints than the harmonic one, regardless the number of spectroscopic bins used for the latter. However, the harmonic approach with 40 bins produces a slightly higher FoM than the hybrid one in the optimistic scenario.

Fig. 4.20 shows that the hybrid approach performs drastically better in constraining h, producing a 1–σ uncertainty on it which is always more than 70% smaller than the one in the harmonic approach. The hybrid approach gives better uncertainties than the harmonic one for n_s and $\Omega_{b,0}$ too. For n_s the uncertainty of the hybrid approach is always smaller than the one of the harmonic approach by 30-40%. For $\Omega_{b,0}$ the hybrid approach gives a 60% smaller uncertainty than the harmonic approach with 4 bins, while the difference is about

20% with 40 bins. Concerning the uncertainty on $\Omega_{\mathrm{m},0}$ the two approaches produce results that are always comparable within 5%. Finally, on σ_8 the harmonic approach with 12 bins performs slightly better than the hybrid one. In the optimistic scenario the uncertainties on σ_8 are always comparable, while in the pessimistic case the harmonic approach produces a 10% smaller uncertainty when using 40 tomographic bins.

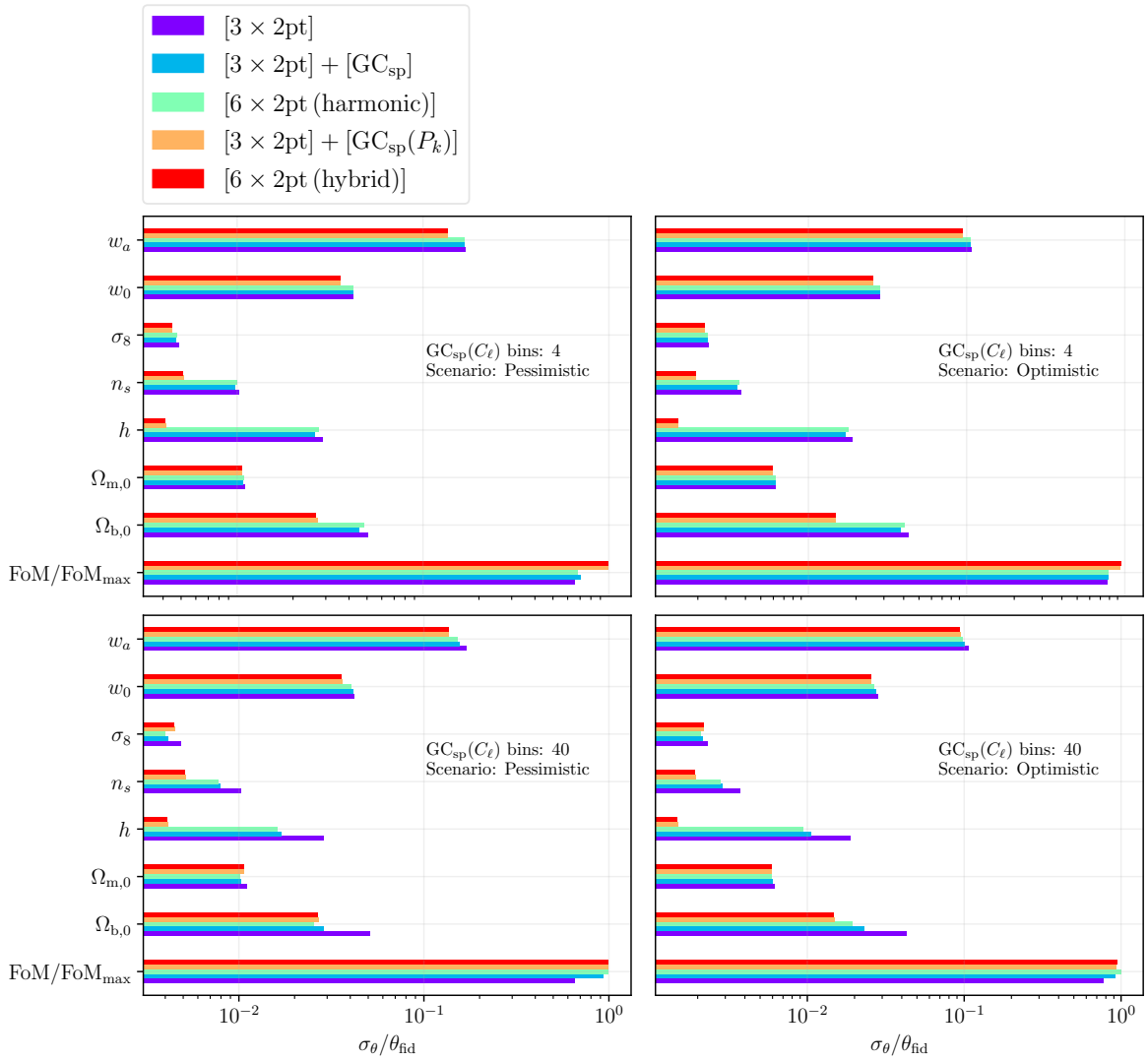

Figure 4.21: Comparison between harmonic and hybrid approaches for the case of 6×2pt statistics. The top panels refer to the optimistic scenario, while the bottom panels refer to the pessimistic one.

4.4 Conclusions and outlook

In this work I presented the results of the first *Euclid* forecasts which include all the pairwise correlations between the main probes: weak lensing (WL), photometric galaxy clustering

($\mathrm{GC_{ph}}$), and spectroscopic galaxy clustering ($\mathrm{GC_{sp}}$). I have studied these correlations from two points of view. First, as off-diagonal terms of the covariance matrix, the *cross-covariances*, that account for the statistical correlation between two different probes. Second, as additional observables, the *cross-correlation signals* which can be sensitive to cosmological parameters.

My work is a natural extension of the *Euclid* IST forecast [5], where the authors studied WL and $\mathrm{GC_{ph}}$ including their correlation in the 2D harmonic domain. The $\mathrm{GC_{sp}}$ was instead studied in the 3D Fourier domain, and in first approximation was assumed to be independent from the other two probes. Here I extended the harmonic formalism also to $\mathrm{GC_{sp}}$, with the aim of including the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ and $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$ correlations in the analysis as well.

I have performed the forecasts in two scenarios: *optimistic* and *pessimistic*, differing only for the multipole ranges chosen for the harmonic power spectra. In the optimistic scenario, the multipole range was $10 \leq \ell \leq 3000$ for both $\mathrm{GC_{ph}}$ and the projected $\mathrm{GC_{sp}}$ and $10 \leq \ell \leq 5000$ for WL. In the pessimistic scenario the ranges were $10 \leq \ell \leq 750$ and $10 \leq \ell \leq 1500$, respectively.

I have considered two different approaches to include the cross-correlations in the analysis: the *harmonic* approach and the *hybrid* approach. In the harmonic approach all the observables – the two-point correlation functions – are treated in the harmonic domain, i.e. using the $C(\ell)$'s formalism. This approach allows us to naturally take into account all cross-covariances between the observables, computed via Eq. (4.28). Nonetheless, it has the disadvantage to significantly lower the constraining power of $\mathrm{GC_{sp}}$, since the integral along the line of sight prevents to fully exploit the accurate radial information provided by the spectroscopic clustering. In order to recover such information, I tried to refine the tomographic binning of the projected $\mathrm{GC_{sp}}$, from 4 bins (the baseline setting) up to a maximum of 40 bins.

In the hybrid approach all the observables are studied in the harmonic domain – including $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$ and $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ – except for the $\mathrm{GC_{sp}}$ auto-correlation function. I have considered it as an independent observable, adding the $\mathrm{GC_{sp}}$ Fisher matrix that had been computed in IST:F [5], using the 3D Fourier power spectrum as observable. The main advantage of this approach is that it fully exploits the potential of $\mathrm{GC_{sp}}$, keeping the information from radial BAO and RSD available thanks to accurate spectroscopic redshift measurements. At the same time the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ and $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$ cross-correlations are included in the analysis as harmonic two-point functions, i.e. $C(\ell)$'s. However, this approach comes with two main drawbacks. First, for consistency with the official 3D spectroscopic galaxy clustering setting, only 4 tomographic bins can be employed for computing the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ and $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$ cross-correlations, thus limiting the gain of constraining power which comes from their inclusion in the analysis. Second, it is not obvious how to compute the cross-covariance terms between the 3D $\mathrm{GC_{sp}}$ auto-correlation and the other 2D observables. Therefore, accordingly to the conclusions from the analysis in the harmonic domain (which correctly accounts for the projected part of such cross-covariances), in the hybrid approach they have been neglected, leaving their computation to future work.

Here I summarise the results of the analysis in the three cases considered: the combination of $\mathrm{GC_{ph}}$ and $\mathrm{GC_{sp}}$, the combination of WL and $\mathrm{GC_{sp}}$, and the so-called $6 \times 2\mathrm{pt}$, i.e. the full combination of $\mathrm{WL}, \mathrm{GC_{ph}}, \mathrm{GC_{sp}}$ altogether. In the latter case – which is the most interesting for the *Euclid* data analysis – I presented the results in both the optimistic and pessimistic scenarios. Instead, for the two pairwise combinations $\mathrm{GC_{ph}}$-$\mathrm{GC_{sp}}$ and WL-$\mathrm{GC_{sp}}$, I reported the results for the optimistic scenario alone.

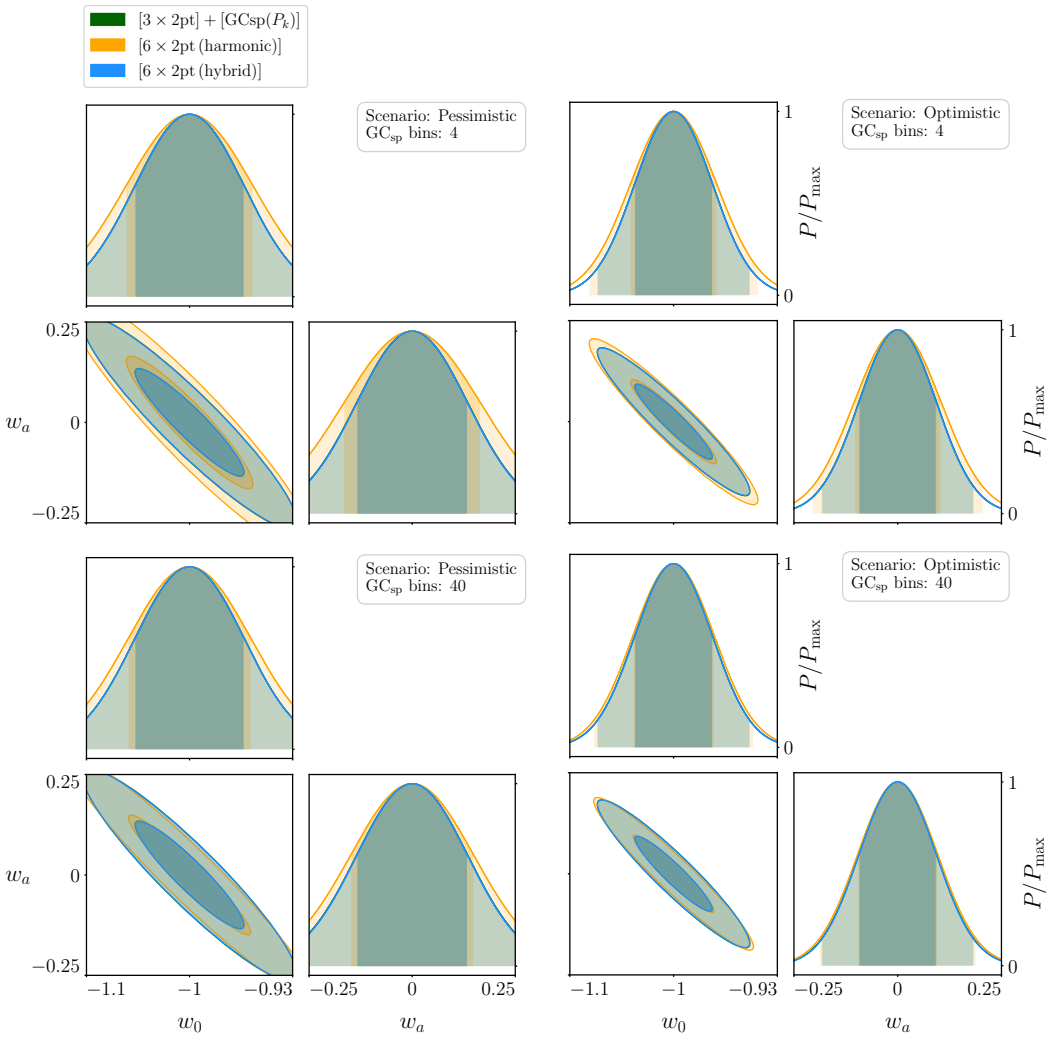

Figure 4.22: Contour plots showing the comparison between the two approaches in the case of the 6×2pt statistics.

Combination of $\mathbf{GC_{ph}}$ and $\mathbf{GC_{sp}}$ In the harmonic approach, the full analysis, $[\mathrm{GC_{ph}} + \mathrm{GC_{sp}} + \mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})]$, i.e. including cross-covariance and cross-spectrum, provides a FoM of 69 in the baseline setting of 4 spectroscopic bins. In the combination of $\mathrm{GC_{ph}}$ and $\mathrm{GC_{sp}}$, the cross-covariance between the two (as said above, computed only in the harmonic approach) can be considered practically negligible, as it does not change the uncertainties on the cosmological parameters by more than 6% and the FoM by more than 3%, with respect to the reference combination in the harmonic domain, $[\mathrm{GC_{ph}}]+[\mathrm{GC_{sp}}]$, of the two probes taken as independent. In the harmonic approach, the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ cross-correlation signal starts to be significant only when using 40 bins, yielding a maximum FoM of 153, which is 38% higher than for the $[\mathrm{GC_{ph}}] + [\mathrm{GC_{sp}}]$ combination. Regarding the uncertainties on w_0 and w_a, the improvement given by the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ cross-correlation is 5% and 6% respectively in the baseline 4 bins setting, while it is 23% and 25% when using 40 bins. The gain on the uncertainties on the other parameters is less than 5% with 4 bins and ranges from 10% to 20% with 40 bins.

In the hybrid approach, the full analysis, $[\mathrm{GC_{ph}} + \mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})] + [\mathrm{GC_{sp}}(P_k)]$, including cross-spectrum but, as said above, neglecting cross-covariance, gives instead a FoM of 234, which is much higher than in the harmonic case. However, in the hybrid case the impact of the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ signal is negligible, as combining the probes as independent, $[\mathrm{GC_{ph}}] + [\mathrm{GC_{sp}}(P_k)]$, provides a FoM of 230, which is only 1.7% lower than for the full analysis. The gain on the marginalised uncertainties is even smaller, being always less than 1%. Therefore, it is possible to conclude that the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ cross-correlation can be neglected in the combination of $\mathrm{GC_{ph}}$ with $\mathrm{GC_{sp}}$.

Combination of WL and $\mathbf{GC_{sp}}$ In the harmonic approach, the full analysis, $[\mathrm{WL} + \mathrm{GC_{sp}} + \mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})]$, i.e. including cross-covariance and cross-spectrum, provides a FoM of 103 in the baseline setting of 4 spectroscopic bins. In the combination of WL and $\mathrm{GC_{sp}}$, the cross-covariance (computed only in the harmonic approach) is even more negligible than for the combination of $\mathrm{GC_{ph}}$ with $\mathrm{GC_{sp}}$, impacting the constraints always by less than 1% with respect to the reference independent combination, $[\mathrm{WL}] + [\mathrm{GC_{sp}}]$, computed in the harmonic domain. In this domain, the $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$ cross-correlation signal improves the constraints almost independently of the number of tomographic bins used for $\mathrm{GC_{sp}}$, when compared to $[\mathrm{WL}]+[\mathrm{GC_{sp}}]$. The percentage gain on the FoM produced by the XC inclusion is 38% and 33% with 4 and 40 bins respectively. This counterintuitive trend has been explained in Sec. 4.3.3: it is due to the fact that the performance of the harmonic $\mathrm{GC_{sp}}$ auto-spectrum alone improves faster than for $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$ when refining the spectroscopic binning. The value of the FoM provided by the full combination, $[\mathrm{WL} + \mathrm{GC_{sp}} + \mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})]$, increases to 188 with 40 bins, i.e. by $\sim 45\%$ with respect to the baseline bin setting. The $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$ improvement on the marginalised uncertainties on w_0 and w_a exhibits the same trend observed for the FoM, decreasing from 20% to 13% as the number of spectroscopic bins increases from 4 to 40. This happens also for the uncertainties on $\Omega_{\mathrm{m},0}$ and σ_8, with improvements ranging from 18% to 12% for the former and from 16% to 11% for the latter. The decrease on the uncertainties on $\Omega_{\mathrm{b},0}$ and n_s ranges instead from 5% to 8% for the former and from 5% to 2% on the latter. Finally, the uncertainty on h is the only exception to this trend, as its decrease improves from 1% to 4%.

In the hybrid approach, the full analysis, $[\mathrm{WL} + \mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})] + [\mathrm{GC_{sp}}(P_k)]$, including cross-spectrum but neglecting cross-covariance, provides a FoM of 183, which is comparable to the harmonic case when 40 bins are used, but is $\sim 56\%$ larger in the case of 4 bins. In the

hybrid approach, the impact of the $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$ signal is not negligible, since the FoM is 15% larger than for the independent combination, $[\mathrm{WL}] + [\mathrm{GC_{sp}}(P_k)]$, of the two probes. The improvements on the marginalised uncertainties are all comprised between 4% and 8%, with no significant differences between the parameters. The uncertainty on $\Omega_{\mathrm{m},0}$ is the one improving the most, about 8%, while the one gaining the least is the uncertainty on n_{s}, decreasing by slightly less than 5%.

Combination of $\mathbf{WL}, \mathbf{GC_{ph}}, \mathbf{GC_{sp}}$ and the $\mathbf{6 \times 2pt}$ statistics In the harmonic approach, the full 6×2pt analysis, Eq. (4.58), i.e. including all cross-covariances and cross-spectra, provides a FoM of 1018 in the baseline setting of 4 spectroscopic bins. The overall impact of the cross-covariances (computed only in the harmonic approach) between the angular $\mathrm{GC_{sp}}$ and the observables of the 3×2pt statistics, with respect to the independent combination $[3 \times 2\mathrm{pt}] + [\mathrm{GC_{sp}}]$, is slightly higher than for the pairwise combinations $\mathrm{GC_{ph}}$-$\mathrm{GC_{sp}}$ and WL-$\mathrm{GC_{sp}}$, reported above. This is presumably due to the cumulative effect of three independent off-diagonal covariance blocks. These account for the covariances between the $\mathrm{GC_{sp}}$ auto-spectrum and the three observable power spectra of the 3×2pt statistics, namely the auto-spectra of WL, $\mathrm{GC_{ph}}$ and their cross-spectrum $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{ph}})$. However, the increase on parameter uncertainties is never larger than 8%, as Fig. 4.16 shows.

Both in the harmonic and hybrid approaches, the cross-correlations with $\mathrm{GC_{sp}}$ are always negligible. The $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$ is definitely negligible both in the optimistic and pessimistic scenarios. In the harmonic approach, it always improves parameter constraints by less than 0.05%, with respect to the independent combination, $[3 \times 2\mathrm{pt} + \mathrm{GC_{sp}}]$, computed in the same approach. This is because the contribution brought by the $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$ is dominated by the $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{ph}})$, which is already present in the Fisher used as reference. Moreover, still in the harmonic approach, the effect of the $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$ becomes completely invisible with respect to the results from the independent combination, $[3 \times 2\mathrm{pt}] + [\mathrm{GC_{sp}}]$, computed in the same approach. This is because the difference is dominated by the cross-covariance between $\mathrm{GC_{sp}}$ and 3×2pt, which is accounted when adding the cross-correlation observable.

In the hybrid approach the situation is similar: the $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$ cross-correlation improves both the FoM and the uncertainties by always less than 0.05% with respect to the constraints produced by the $[3 \times 2\mathrm{pt}] + [\mathrm{GC_{sp}}(P_k)]$ combination.

The effect of the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ is larger than the one of $\mathrm{XC}(\mathrm{WL}, \mathrm{GC_{sp}})$, both in the optimistic and pessimistic scenarios and both in the harmonic and in the hybrid approach. For this reason the 6×2pt statistics is essentially equivalent to adding $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ only. Nonetheless, the improvement on the constraints produced by the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ cross-correlation is almost always smaller than 10%. In the harmonic approach, its improvement is always below 5% when compared to $[3 \times 2\mathrm{pt} + \mathrm{GC_{sp}}]$, while it is dominated by the covariance effect when compared to $[3 \times 2\mathrm{pt}] + [\mathrm{GC_{sp}}]$. The only exception is when using 40 bins for $\mathrm{GC_{sp}}$: in this case the FoM improves by about 10%, while the marginalised uncertainties on $\Omega_{\mathrm{b},0}$ and h decrease by $\sim 15\%$. The uncertainties on the other parameters decrease by always less than 5% instead.

In the hybrid approach the $\mathrm{XC}(\mathrm{GC_{ph}}, \mathrm{GC_{sp}})$ improvement is always below 2%, for both FoM and uncertainties with respect to $[3 \times 2\mathrm{pt}] + [\mathrm{GC_{sp}}(P_k)]$, as in this case only the baseline setting of 4 spectroscopic bins can be used for consistency with the official 3D spectroscopic galaxy clustering approach.

Finally, looking at the absolute performance, the values of the FoM are reported in Tab. 4.8, which is possible to summarise as:

1. the $[3 \times 2\text{pt}] + [\text{GC}_{\text{sp}}(P_k)]$ is taken as reference, and it gives a FoM of 1216 (549) in the optimistic (pessimistic) scenario;

2. the harmonic $6 \times 2\text{pt}$ statistics with 40 bins gives a FoM of 1296 (550) in the optimistic (pessimistic) scenario, which is only 6.6% (0.26%) better than 1. When using 4 bins it gives a FoM of 1018 (380) in the optimistic (pessimistic) scenario, which instead is 15% (31%) worse than 1.

3. The hybrid $6 \times 2\text{pt}$ statistics is essentially equivalent to the $[3 \times 2\text{pt}] + [\text{GC}_{\text{sp}}(P_k)]$, with a FoM of 1228 (556) in the optimistic (pessimistic) scenario, only 1.18% (0.95%) better than 1.

Therefore, $\text{XC}(\text{GC}_{\text{ph}}, \text{GC}_{\text{sp}})$ and $\text{XC}(\text{WL}, \text{GC}_{\text{sp}})$ cross-correlations are negligible when added to the combination of GC_{sp} and the $3 \times 2\text{pt}$ statistics, both in the harmonic and hybrid approaches.

In conclusion, either the cross-covariances (here computed only in the harmonic approach) or the cross-correlations (computed both in the harmonic and hybrid approaches) between GC_{sp} and the other main probes of *Euclid*, represented by the so-called $3 \times 2\text{pt}$ statistics, have a negligible impact on the cosmological parameter constraints.

In the case of the hybrid approach, I attribute this result to the effect of the $\text{XC}(\text{WL}, \text{GC}_{\text{ph}})$ cross-correlation which is dominant with respect to the other cross-correlations, and to the higher performance of the full anisotropic 3D GC_{sp} probe with respect to the projected one.

In the case of the 2D harmonic domain, I attribute this result to two main theoretical limitations of the projected GC_{sp} probe: the high shot noise and the limited redshift range of the sample. I have found that, under two conditions, GC_{sp} in harmonic space becomes equivalent to GC_{ph} in terms of constraining power, as it can be seen from Fig. 4.7. The first condition is to reduce the shot noise of GC_{sp} to the same level of GC_{ph}. The second is to restrict the tomographic bins of GC_{ph} to the 4 photometric bins comprised in the GC_{sp} redshift range, i.e. $z \in [0.9, 1.8]$. Under these same conditions, $\text{XC}(\text{WL}, \text{GC}_{\text{sp}})$ and $\text{XC}(\text{WL}, \text{GC}_{\text{ph}})$ become equivalent in terms of constraining power too, as I show in Fig. 4.14. Nonetheless, these conditions are not realistic.

Finally, a future extension of my work would be to overcome some approximations made in this thesis. First, I have computed the $C(\ell)$'s making use of the Limber approximation. It has been shown in a recent work [86] that this may result in a biased analysis. Using the exact expression for computing the angular power spectra would help to prevent this issue. Second, I have computed the $C(\ell)$'s covariance with Eq. (4.28) as in IST:F [5], and this formula only accounts for the Gaussian contributions. Comparison with covariances estimated from N-body simulations showed that the inclusion of non-Gaussian effects may be necessary to reach a better agreement with simulations [87]. In particular, including the super-sample covariance effect may help to improve Fisher forecasts by making the signal-to-noise ratio more realistic [88]. Third, in what I have called the "hybrid approach" I have neglected the covariances between the Fourier GC_{sp} auto-correlation and the other observables. The computations done here with GC_{sp} in the harmonic space suggest that these terms may be

negligible. Nonetheless, providing an analytical modelling of these terms when $\mathrm{GC_{sp}}$ is studied in the 3D Fourier space would surely help to confirm this hypothesis.

Conclusions

This thesis work has been carried out in the context of the ESA *Euclid* mission. As a member of the Euclid Consortium, I contributed to two different areas of the mission:

- I worked on spectroscopic image simulations of NISP raw data, in order to validate the data-reduction software pipeline of the spectroscopic channel of *Euclid*;
- I performed a forecast about the inclusion of the correlations between the *Euclid* main probes.

In the first half of my PhD I collaborated with the Milan INAF-IASF OU-SIR group in order to perform the simulations necessary for the validation of the SIR data-reduction Pipeline. I extended the official spectroscopic image simulator of *Euclid* built by OU-SIM, making it more flexible and allowing full control on the simulation side. The most useful features of the software are essentially two. First, the possibility to choose in a simple way the instrumental effects to be simulated, together with the coordinates and the spectra of sources falling in the field-of-view. Second, the possibility to run a full simulation-reduction chain, in which the SIR Pipeline can be executed on the simulated images to extract the spectra. These features of the software allowed the SIR group to use it for producing the input data for various validation test cases: simple debugging of the reduction algorithms, optimising spectra extraction parameters or testing the overall performances of the whole pipeline. At the moment of writing there are also plans for applying it to the production of simulated spectra for the *Euclid* Legacy science groups, in particular the Active Galactic Nuclei (AGN) and Galaxy Evolution working groups.

In the second half of my PhD I collaborated within the *Euclid* GC-SWG WP:Likelihood in order to perform a Fisher forecast about the inclusion of the correlations between the *Euclid* main probes: weak lensing (WL), photometric (GC_{ph}) and spectroscopic (GC_{sp}) galaxy clustering. The aim of the work was to assess the impact of these correlations on the dark energy Figure-of-Merit (FoM) and the marginalised constraints on the cosmological parameters. In a previous forecast – the *Euclid* IST forecast [5], hereafter briefly denoted as IST:F – it was shown that the cross-correlation between WL and GC_{ph} has a significant impact on the scientific outcome of the mission, improving the FoM up to 400% and the constraints on the parameters up to 60%. In particular, WL and GC_{ph} and their cross-correlation were studied in the 2D harmonic space, using tomographic angular power spectra as observables. The GC_{sp} probe was instead studied as independent from the other two, using the 3D Fourier galaxy power spectrum as the observable.

The work presented here extends the IST:F by including also the correlations between GC_{sp} and WL and between GC_{sp} and GC_{ph}. A proper weight function for GC_{sp} is introduced, allowing us to compute the $XC(WL, GC_{sp})$ and $XC(GC_{ph}, GC_{sp})$ cross-correlations in the harmonic

domain, and therefore including them as observables for the Fisher forecast. The $\mathrm{GC_{sp}}$ auto-correlation has been included following two different approaches: the *harmonic* approach and the *hybrid* approach. In the harmonic approach it is implemented as a tomographic angular power spectrum: this allows to naturally take into account the cross-covariances with the other observables, but it significantly limits its constraining power because of the integration along the line of sight. In order to circumvent this issue and partially recover the radial information, I refined the tomographic binning of $\mathrm{GC_{sp}}$, starting from the 4 bins used in the IST:F up to 40 equally spaced bins. In the hybrid approach, the $\mathrm{GC_{sp}}$ auto-correlation is included as an independent probe by adding *a posteriori* its Fisher matrix, separately computed using the the 3D full anisotropic power spectrum, to the $[3\times2\mathrm{pt}+\mathrm{XC}(\mathrm{WL},\mathrm{GC_{sp}})+\mathrm{XC}(\mathrm{GC_{ph}},\mathrm{GC_{sp}})]$ Fisher matrix. In this way the full constraining power of $\mathrm{GC_{sp}}$ is preserved, but the covariances of its auto-correlation with the other probes are neglected.

I have performed the forecast in two scenarios: an *optimistic* scenario and a *pessimistic* one, differing for the multipole values in which the angular power spectra are computed. In the optimistic scenario the multipole range is $10 \leq \ell \leq 3000$ for angular GC and $10 \leq \ell \leq 5000$ for WL, while in the pessimistic scenario it is $10 \leq \ell \leq 750$ for GC and $10 \leq \ell \leq 1500$ for WL. My analysis shows that the $\mathrm{XC}(\mathrm{WL},\mathrm{GC_{sp}})$ and $\mathrm{XC}(\mathrm{GC_{ph}},\mathrm{GC_{sp}})$ cross-correlations have a much minor impact on the results than the $\mathrm{XC}(\mathrm{WL},\mathrm{GC_{ph}})$. Regarding this last, the results agree with the ones obtained by IST:F, confirming its significance in improving the constraints on cosmological parameters.

The cross-covariance terms between the $\mathrm{GC_{sp}}$ auto-correlation and the other observables – computed in the harmonic domain – always have a negligible impact on the constraints, the maximum percentage variation being of 8% in the combination of $\mathrm{GC_{sp}}$ with the $3\times2\mathrm{pt}$ statistics.

The cross-correlation observables do not add significant constraining power too. In the pairwise combination of WL and $\mathrm{GC_{sp}}$ the $\mathrm{XC}(\mathrm{WL},\mathrm{GC_{sp}})$ improves the FoM by about 38% in the harmonic approach and by 15% in the hybrid, which is more than a factor of 10 smaller than the contribution given by its photometric counterpart $\mathrm{XC}(\mathrm{WL},\mathrm{GC_{ph}})$. In the combinations of $3\times2\mathrm{pt}$ statistics with $\mathrm{GC_{sp}}$, when $\mathrm{XC}(\mathrm{WL},\mathrm{GC_{ph}})$ is included as well, the contribution of the $\mathrm{XC}(\mathrm{WL},\mathrm{GC_{sp}})$ cross-correlation becomes completely negligible, impacting the constraints only by 0.05%, both in the optimistic and pessimistic scenarios. The $\mathrm{XC}(\mathrm{GC_{ph}},\mathrm{GC_{sp}})$ is always irrelevant in the hybrid approach, impacting the constraints by always less than 2%. In the harmonic approach it gives a non-negligible contribution only in the pairwise combination of $\mathrm{GC_{ph}}$-$\mathrm{GC_{sp}}$, for which it improves the FoM by 38% when 40 tomographic bins are used for $\mathrm{GC_{sp}}$. Its contribution drops to 5% when the $\mathrm{XC}(\mathrm{WL},\mathrm{GC_{ph}})$ is included in the analysis, either in the optimistic and the pessimistic scenario. Therefore, the $\mathrm{XC}(\mathrm{GC_{ph}},\mathrm{GC_{sp}})$ and $\mathrm{XC}(\mathrm{WL},\mathrm{GC_{sp}})$ cross-correlations are essentially negligible when added to the combination of $\mathrm{GC_{sp}}$ and the $3\times2\mathrm{pt}$ statistics, both in the harmonic and hybrid approaches.

The analysis shows that the hybrid approach performs better than the harmonic. This is due to the loss of information in the angular projection of the $\mathrm{GC_{sp}}$. Refining the tomographic binning partially helps to mitigate this loss of radial information.

The main theoretical limitations of $\mathrm{GC_{sp}}$ in the harmonic domain, when compared to $\mathrm{GC_{ph}}$, are the higher shot noise and the reduced redshift range. I have in fact shown that the two probes become comparable to each other under two conditions. First, the shot noise of $\mathrm{GC_{sp}}$ has to be reduced to the same level of $\mathrm{GC_{ph}}$. Second, only the tomographic bins of $\mathrm{GC_{ph}}$ falling within the redshift range of $\mathrm{GC_{sp}}$ have to be considered. As it has been already

explained in Sec. 4.4, these conditions are not realistic, and the aforementioned limitations of the harmonic GC_{sp} probe cannot be overcome in this way.

I therefore conclude that probably it will be possible to neglect the GC_{sp}-WL and GC_{sp}-GC_{ph} correlations in the future *Euclid* data analysis. Obviously this is not the last word on the subject, as in the next future this forecast will be improved under various aspects. The super-sample covariance contribution will be added to the covariance matrix, providing a more realistic signal-to-noise ratio [87, 88]. The Limber approximation will be dropped in the computation of the angular power spectra, in order to prevent eventual biased results [86]. Finally, the cross-covariance terms between 2D and 3D power spectra – here neglected in the so-called hybrid approach – will be analytically modelled and computed.

Appendix A

Fisher matrix expressions for gaussian fields

A.1 Introduction

Under the assumption that the probes can be represented as a collection of correlated gaussian fields $\boldsymbol{a}_{\ell m}$, there exists a closed expression for the Fisher matrix elements. In [5] two expressions are reported, referred as *second-order covariance* and *fourth-order covariance*. Here I provide a proof for the second-order expression, and I show that it is equivalent to the fourth-order expression, following the treatment of [89].

Second-order covariance formula

For the auto-correlation of a probe A the Fisher matrix element reads

$$F^{AA}_{\alpha\beta} = \frac{1}{2}\sum_{\ell=\ell_{\rm min}}^{\ell_{\rm max}} (2\ell+1)\,\mathrm{Tr}\left[\frac{\partial \mathbf{C}^{AA}(\ell)}{\partial\alpha}\left(\mathbf{\Sigma}^{AA}(\ell)\right)^{-1}\frac{\partial \mathbf{C}^{AA}(\ell)}{\partial\beta}\left(\mathbf{\Sigma}^{AA}(\ell)\right)^{-1}\right]. \tag{A.1}$$

In the case of the cross-correlation between $\mathcal{N}$ probes the formula is the following [5]

$$F^{\rm XC}_{\alpha\beta} = \frac{1}{2}\sum_{\ell=\ell_{\rm min}}^{\ell_{\rm max}} (2\ell+1)\sum_{ABCD}\sum_{ijkm}\frac{\partial C^{AB}_{ij}(\ell)}{\partial\alpha}\left[\left(\Sigma^{BC}(\ell)\right)^{-1}\right]_{jk}\frac{\partial C^{CD}_{km}(\ell)}{\partial\beta}\left[\left(\Sigma^{DA}(\ell)\right)^{-1}\right]_{mi}. \tag{A.2}$$

The sum runs over all values of the probe indices A, B, C, D and the tomographic indices i, j, k, m, and $\left[\mathbf{\Sigma}^{AB}(\ell)\right]_{ij} = \Sigma^{AB}_{ij}(\ell)$. This equation can be recast in the same form of the auto-correlation. First, define the following multi-indices

$$I \equiv (A, i) \qquad J \equiv (B, j) \qquad K \equiv (C, k) \qquad M \equiv (D, m)\,, \tag{A.3}$$

with significance increasing from right to left. Now define a block matrix $\mathbf{C}^{XC}(\ell)$ such that:

$$\mathbf{C}^{XC}(\ell) \equiv \begin{pmatrix} \mathbf{C}^{11}(\ell) & \mathbf{C}^{12}(\ell) & \dots & \mathbf{C}^{1\mathcal{N}}(\ell) \\ \left[\mathbf{C}^{12}(\ell)\right]^T & \mathbf{C}^{22}(\ell) & \dots & \mathbf{C}^{2\mathcal{N}}(\ell) \\ \vdots & \vdots & \ddots & \vdots \\ \left[\mathbf{C}^{1\mathcal{N}}(\ell)\right]^T & \left[\mathbf{C}^{2\mathcal{N}}(\ell)\right]^T & \dots & \mathbf{C}^{\mathcal{N}\mathcal{N}}(\ell) \end{pmatrix}, \qquad \left[\mathbf{C}^{XC}(\ell)\right]_{IJ} = C^{AB}_{ij}(\ell)\,. \tag{A.4}$$

This block matrix is symmetric by construction:

$$\left[\mathbf{C}^{XC}(\ell)\right]^T = \mathbf{C}^{XC}(\ell)\,.$$

A block matrix $\mathbf{\Sigma}^{XC}(\ell)$ can be defined with the same conventions. Therefore, the nested sum of Eq. (A.2) can be rewritten as

$$\sum_{IJKM} \left(\frac{\partial \mathbf{C}^{XC}(\ell)}{\partial\alpha}\right)_{IJ} \left[\left(\mathbf{\Sigma}^{XC}(\ell)\right)^{-1}\right]_{JK} \left(\frac{\partial \mathbf{C}^{XC}(\ell)}{\partial\beta}\right)_{KM} \left[\left(\mathbf{\Sigma}^{XC}(\ell)\right)^{-1}\right]_{MI},$$

which is explicitly a matrix trace:

$$F^{XC}_{\alpha\beta} = \frac{1}{2}\sum_{\ell}(2\ell+1)\,\mathrm{Tr}\left[\frac{\partial \mathbf{C}^{XC}(\ell)}{\partial\alpha}\left(\mathbf{\Sigma}^{XC}(\ell)\right)^{-1}\frac{\partial \mathbf{C}^{XC}(\ell)}{\partial\beta}\left(\mathbf{\Sigma}^{XC}(\ell)\right)^{-1}\right]. \tag{A.5}$$

Fourth-order covariance formula

The fourth order covariance formula for the Fisher matrix of the AA auto-correlation reads

$$F^{AA}_{\alpha\beta} = \sum_{\ell=\ell_{\min}}^{\ell_{\max}} \sum_{i\le j,k\le m} \frac{\partial C^{AA}_{ij}(\ell)}{\partial\theta_\alpha}\,\mathrm{Cov}^{-1}\left[C^{AA}_{ij}(\ell), C^{AA}_{km}(\ell)\right]\frac{\partial C^{AA}_{km}(\ell)}{\partial\theta_\beta}\,, \tag{A.6}$$

while when including also cross-correlations there is also a sum over the block descriptors A, B, C, D

$$F^{\mathrm{XC}}_{\alpha\beta} = \sum_{\ell=\ell_{\min}}^{\ell_{\max}} \sum_{(AB,CD)_{\mathrm{sym}}} \sum_{(ij,km)_{\mathrm{sym}}} \frac{\partial C^{AB}_{ij}(\ell)}{\partial\theta_\alpha}\,\mathrm{Cov}^{-1}\left[C^{AB}_{ij}(\ell), C^{CD}_{km}(\ell)\right]\frac{\partial C^{CD}_{km}(\ell)}{\partial\theta_\beta}\,. \tag{A.7}$$

The notation $(ij,km)_{\mathrm{sym}}$ means that the sum runs run over the pairs ij, km such that $i \le j$ when $A = B$ and $k \le m$ when $C = D$. Analogously $(AB,CD)_{\mathrm{sym}}$ stands for summing over the unique pairs AB, CD independently of the order of the letters A, B and C, D in the single pairs. From a combinatorical point of view this means the pairs AB, CD are the combinations with replacement of two elements drawn from $\mathcal{N}$, i.e. the 2-combinations of $\mathcal{N}$ elements. This means that for the auto-correlation only the upper triangular part of the $C(\ell)$ matrices is to be considered. To make things clearer Eq. (A.7) can be turned into the same form of Eq. (A.6), as it has been done in the previous section. In particular

$$\sum_{(AB,CD)_{\mathrm{sym}}} \sum_{(ij,km)_{\mathrm{sym}}} (\dots) = \sum_{I\le J,K\le M} (\dots)\,.$$

Considering only the upper triangular part of the block matrix Eq. (A.4) precisely means to take only the upper triangular part of the auto-correlations and the whole matrices for all the 2-combinations of pairs $A \neq B, C \neq D$. Thus

$$F^{\mathrm{XC}}_{\alpha\beta} = \sum_{\ell=\ell_{\min}}^{\ell_{\max}} \sum_{I\le J,K\le M} \frac{\partial C^{XC}_{IJ}(\ell)}{\partial\theta_\alpha}\,\mathrm{Cov}^{-1}\left[C^{XC}_{IJ}(\ell), C^{XC}_{KM}(\ell)\right]\frac{\partial C^{XC}_{KM}(\ell)}{\partial\theta_\beta}\,. \tag{A.8}$$

A.2 Linear algebra tools

In this section I follow [83], referring to it for all missing proofs.

Matrix vectorisation The vectorisation of a $p \times q$ matrix X is defined as an $pq \times 1$ vector

$$\mathrm{vec}(X) = \begin{pmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_m \end{pmatrix} , \tag{A.9}$$

where $\boldsymbol{x}_i$ is the i-th column of X.

Commutation matrix The commutation matrix K_{pq} of order $pq \times pq$ is defined as

$$K_{pq} = \sum_{i=1}^{p} \sum_{j=1}^{q} \left(H_{ij} \otimes H_{ij}^T \right) , \tag{A.10}$$

where the $p \times q$ matrix H_{ij} has a unit element at $(i,j)^{\mathrm{th}}$ position and zero elsewhere. The commutation matrix has the following property

$$K_{pq}(A \otimes B)K_{rs} = B \otimes A \,, \tag{A.11}$$

where A, B have sizes $p \times q$ and $r \times s$ respectively. The symbol $\otimes$ is the Kronecker product of matrices: here I will simply use its properties, referring to [83] for the proofs. If X is $p \times p$ square matrix then

$$K_{pp}(X \otimes X) = (X \otimes X)K_{pp} \,. \tag{A.12}$$

The typical element of K_{pp} is

$$(K_{pp})_{ij,km} = \delta_{im}\delta_{jk} \,. \tag{A.13}$$

Half-vectorisation If X is a $p \times p$ symmetric matrix the half-vectorisation of X is defined as a $p(p+1)/2$ vector formed from the elements of the upper triangular sector of the matrix, taken columnwise

$$\mathrm{vecp}(X) = \begin{pmatrix} x_{11} \\ x_{12} \\ x_{22} \\ \vdots \\ x_{1p} \\ \vdots \\ x_{pp} \end{pmatrix} = \mathrm{vecp}\left(X^T\right) . \tag{A.14}$$

Transition matrix The half-vectorisation is related to the vectorisation through the transition matrix B_p, which has size $p^2 \times p(p+1)/2$ and is such that

$$\begin{aligned} \mathrm{vecp}(X) &= B_p^T \,\mathrm{vecp}(X) \\ \mathrm{vec}(X) &= (B_p^+)^T \,\mathrm{vecp}(X) \,, \end{aligned} \tag{A.15}$$

where B_p^+ has size $p(p+1)/2 \times p^2$ and is the Moore-Penrose *left* pseudo-inverse of B_p

$$B_p^+ \equiv (B_p^T B_p)^{-1} B_p^T \,. \tag{A.16}$$

The typical element of the transition matrix is

$$(B_p)_{ij,km} = \frac{1}{2}\left(\delta_{ik}\delta_{jm} + \delta_{im}\delta_{jk}\right), \quad i \le p, j \le p, k \le m \le p \,. \tag{A.17}$$

It can be shown [83] that B_p has full column rank

$$\mathrm{rank}(B_p) = \frac{p(p+1)}{2} \,, \tag{A.18}$$

and

$$B_p^+ B_p = \mathbb{1}_{\frac{p(p+1)}{2}}, \qquad\qquad B_p B_p^+ = \frac{1}{2}\left(\mathbb{1}_{p^2} + K_{pp}\right) \equiv \mathcal{M}_p \,. \tag{A.19}$$

The $p^2 \times p^2$ matrix $\mathcal{M}_p$ has the typical element

$$(\mathcal{M}_p)_{ij,km} = \frac{1}{2}\left(\delta_{ik}\delta_{jm} + \delta_{im}\delta_{jk}\right), \quad i,j,k,m \le p \,, \tag{A.20}$$

and it satisfies

$$\mathcal{M}_p(X \otimes X) = (X \otimes X)\mathcal{M}_p \,. \tag{A.21}$$

Moreover, the $\mathcal{M}_p$ matrix is singular:

$$\mathrm{rank}(\mathcal{M}_p) = \mathrm{rank}(B_p B_p^+) \le \min\left[\mathrm{rank}(B_p), \mathrm{rank}(B_p^+)\right] \le \mathrm{rank}(B_p) = \frac{p(p+1)}{2} < p^2 \,.$$

Since $\mathcal{M}_p$ has not full rank, it must be $\det(\mathcal{M}_p) = 0$.

Given a $p \times p$ matrix Y, it follows that $Y \otimes Y$ is $p^2 \times p^2$ and its indices can be represented as pairs of the indices of Y:

$$(Y \otimes Y)_{ij,km} = Y_{ik} Y_{jm} \,.$$

Now consider the $p(p+1)/2 \times p(p+1)/2$ matrix $B_p^T (Y \otimes Y) B_p$, with typical element

$$\left[B_p^T (Y \otimes Y) B_p\right]_{ij,km} = \frac{1}{2}\left[Y_{ik}Y_{jm} + Y_{im}Y_{jk}\right], \quad i \le j \le p,\ k \le m \le p \,. \tag{A.22}$$

The determinant of this matrix is

$$\det\left(B_p^T (Y \otimes Y) B_p\right) = 2^{-\frac{1}{2}p(p-1)} \det(Y)^{p+1} \,, \tag{A.23}$$

and, if Y is non-singular

$$\left[B_p^T (Y \otimes Y) B_p\right]^{-1} = B_p^+ (Y^{-1} \otimes Y^{-1})(B_p^+)^T \,. \tag{A.24}$$

On the other hand the $p^2 \times p^2$ matrix $B_p B_p^+ (Y \otimes Y)$ has the same typical element of the previous one:

$$\begin{aligned}\left[B_p B_p^+ (Y \otimes Y)\right]_{ij,km} = \left[\mathcal{M}_p (Y \otimes Y)\right]_{ij,km} &= \frac{1}{2}\left(\delta_{ia}\delta_{jb} + \delta_{ib}\delta_{ja}\right)(Y \otimes Y)_{ab,km} \\ &= \frac{1}{2}\left(Y_{ik}Y_{jm} + Y_{im}Y_{jk}\right), \quad i,j,k,m \le p \,.\end{aligned}$$

But it is singular, since $\det(\mathcal{M}_p) = 0$.

Trace-Kronecker identity A useful matrix identity [83] is

$$\mathrm{Tr}\left(A^T B C D^T\right) = \mathrm{vec}(A)^T \left(D \otimes B\right) \mathrm{vec}(C)\,. \tag{A.25}$$

A special case of this equality follows by assuming all matrices to be symmetric and setting $B = D$

$$\mathrm{Tr}\left(ABCB\right) = \mathrm{vec}(A)^T \left(B \otimes B\right) \mathrm{vec}(C)\,. \tag{A.26}$$

A.3 Equivalence between second order and fourth order Fisher formulas

Here I prove the equivalence between the second-order and fourth-order covariance formulas for the Fisher matrix. In practice, to make use of the fourth-order formula the $C(\ell)$ matrices will be transformed into vectors, since the covariance defined Eq. (4.28) is a matrix labelled by pairs of indices. In other words it is convenient to make use of the matrix vectorisation formalism introduced in section A.2. Here for simplicity I drop the superscripts from the $C(\ell)$ and $\Sigma(\ell)$ matrices, since the results I prove will be generalised for both the auto- and the cross-correlation cases. Therefore I assume to deal with matrices $\mathbf{C}(\ell)$ and $\mathbf{\Sigma}(\ell)$ that are symmetric with shape $p \times p$. Now by using Eq. (A.26) and proceeding as in [89]

$$\begin{aligned}
\mathrm{Tr}\left(\mathbf{A}\mathbf{\Sigma}(\ell)^{-1}\mathbf{D}\mathbf{\Sigma}(\ell)^{-1}\right) &= \mathrm{vec}(\mathbf{A})^T \left(\mathbf{\Sigma}(\ell)^{-1} \otimes \mathbf{\Sigma}(\ell)^{-1}\right) \mathrm{vec}(\mathbf{D}) \\
&= \mathrm{vecp}(\mathbf{A})^T B_p^+ \left(\mathbf{\Sigma}(\ell)^{-1} \otimes \mathbf{\Sigma}(\ell)^{-1}\right) \left(B_p^+\right)^T \mathrm{vecp}(\mathbf{D}) \\
&= \mathrm{vecp}(\mathbf{A})^T \left[B_p^T \left(\mathbf{\Sigma}(\ell) \otimes \mathbf{\Sigma}(\ell)\right) B_p\right]^{-1} \mathrm{vecp}(\mathbf{D}) \\
&= \frac{2}{2\ell+1}\, \mathrm{vecp}(\mathbf{A})^T \left[\frac{2}{2\ell+1} B_p^T \left(\mathbf{\Sigma}(\ell) \otimes \mathbf{\Sigma}(\ell)\right) B_p\right]^{-1} \mathrm{vecp}(\mathbf{D}) \\
&= \mathrm{vecp}(\mathbf{A})^T\, \mathrm{Cov}^{-1}\left[\mathbf{C}(\ell), \mathbf{C}(\ell)\right] \mathrm{vecp}(\mathbf{D})\,,
\end{aligned} \tag{A.27}$$

where the covariance matrix $\mathrm{Cov}\left[\mathbf{C}(\ell), \mathbf{C}(\ell)\right]$ has been defined as

$$\mathrm{Cov}\left[\mathbf{C}(\ell), \mathbf{C}(\ell)\right] \equiv \frac{2}{2\ell+1} B_p^T \left(\mathbf{\Sigma}(\ell) \otimes \mathbf{\Sigma}(\ell)\right) B_p\,. \tag{A.28}$$

Then it follows that

$$\frac{2\ell+1}{2}\, \mathrm{Tr}\left[\frac{\partial \mathbf{C}(\ell)}{\partial \theta_\alpha} \mathbf{\Sigma}(\ell)^{-1} \frac{\partial \mathbf{C}(\ell)}{\partial \theta_\beta} \mathbf{\Sigma}(\ell)^{-1}\right] = \mathrm{vecp}\left(\frac{\partial \mathbf{C}(\ell)}{\partial \theta_\alpha}\right)^T \left[\mathrm{Cov}\left(\mathbf{C}(\ell), \mathbf{C}(\ell)\right)\right]^{-1} \mathrm{vecp}\left(\frac{\partial \mathbf{C}(\ell)}{\partial \theta_\beta}\right)\,. \tag{A.29}$$

Note that using Eq. (A.22) it follows

$$\left[\mathrm{Cov}\left(\mathbf{C}(\ell), \mathbf{C}(\ell)\right)\right]_{ij,km} = \frac{1}{2\ell+1}\left[C_{ik}(\ell) C_{jm}(\ell) + C_{im}(\ell) C_{jk}(\ell)\right], \quad i \le j \le p,\ k \le m \le p\,, \tag{A.30}$$

which is the same form of Eq. (4.28). Now let X, Y be two square matrices of shape p and M a square matrix of shape $p(p+1)/2$. Consider the following sum, in which the ranges of the

indices are explicitly written:

$$\sum_{a\leq b,c\leq d} X_{ab}\, M_{ab,cd}\, Y_{cd} \equiv \sum_{b,c=1}^{p} \sum_{a\leq b,\, c\leq d} X_{ab}\, M_{ab,cd}\, Y_{cd}\,. \tag{A.31}$$

The composite indices $ab \equiv (a,b), cd \equiv (c,d)$ both spans the range $1,\ldots,p(p+1)/2$. Therefore

$$\begin{aligned}\sum_{b,c=1}^{p} \sum_{a\leq b,\, c\leq d} X_{ab}\, M_{ab,cd}\, Y_{cd} &= \sum_{ab=1}^{\frac{p(p+1)}{2}} \sum_{cd=1}^{\frac{p(p+1)}{2}} [\mathrm{vecp}(X)]_{ab}\, [M]_{ab,cd}\, [\mathrm{vecp}(Y)]_{cd} \\ &= \mathrm{vecp}(X)^T M\, \mathrm{vecp}(Y)\,.\end{aligned} \tag{A.32}$$

Finally

$$\sum_{a\leq b,c\leq d} X_{ab}\, M_{ab,cd}\, Y_{cd} = \mathrm{vecp}(X)^T M\, \mathrm{vecp}(Y)\,. \tag{A.33}$$

Therefore the RHS of the fourth order formulas can be rewritten using the half-vectorisation. For the auto-correlation it follows that

$$F^{\mathrm{AA}}_{\alpha\beta} = \sum_{\ell=\ell_{\min}}^{\ell_{\max}} \mathrm{vecp}\left(\frac{\partial \mathbf{C}^{AA}(\ell)}{\partial\theta_\alpha}\right) \left[\mathrm{Cov}\left(\mathbf{C}^{AA}(\ell), \mathbf{C}^{AA}(\ell)\right)\right]^{-1} \mathrm{vecp}\left(\frac{\partial \mathbf{C}^{AA}(\ell)}{\partial\theta_\beta}\right), \tag{A.34}$$

and using Eq. (A.29) the equality between Eq. (A.1) and Eq. (A.6) follows. For the cross-correlation it is convenient to take the rewritten form Eq. (A.8), from which Eq. (A.33) can be directly used to get

$$F^{\mathrm{XC}}_{\alpha\beta} = \sum_{\ell=\ell_{\min}}^{\ell_{\max}} \mathrm{vecp}\left(\frac{\partial \mathbf{C}^{XC}(\ell)}{\partial\theta_\alpha}\right) \left[\mathrm{Cov}\left(\mathbf{C}^{XC}(\ell), \mathbf{C}^{XC}(\ell)\right)\right]^{-1} \mathrm{vecp}\left(\frac{\partial \mathbf{C}^{XC}(\ell)}{\partial\theta_\beta}\right). \tag{A.35}$$

Finally using again Eq. (A.29) the equivalence between Eq. (A.5) and Eq. (A.7) follows.

A.4 Proof of second order covariance formula

A.4.1 The field perspective

The starting hypothesis is that the spherical harmonics expansion coefficients $a^{Ai}_{\ell m}$ follow a multivariate gaussian distribution with zero mean and covariance matrix $\mathbf{S}$

$$L(\boldsymbol{a}_\ell|\boldsymbol{\theta}) = \frac{\exp\left(-\frac{1}{2}\boldsymbol{a}_\ell^{\mathrm{T}}\mathbf{S}^{-1}\boldsymbol{a}_\ell\right)}{\sqrt{(2\pi)^n|\mathbf{S}|}}\,, \tag{A.36}$$

where $\boldsymbol{\theta}$ is the vector of the parameters and $\boldsymbol{a}_\ell$ is a vector collecting the $a^{Ai}_{\ell m}$. The logarithm of this probability distribution function is

$$\ln L(\boldsymbol{a}_\ell|\boldsymbol{\theta}) = -\frac{1}{2}\left[n\ln(2\pi) + \ln|\mathbf{S}| + \boldsymbol{a}_\ell^{\mathrm{T}}\mathbf{S}^{-1}\boldsymbol{a}_\ell\right] = -\frac{1}{2}\left[n\ln(2\pi) + \ln|\mathbf{S}| + \mathrm{Tr}(\mathbf{S}^{-1}\mathbf{A})\right]\,, \tag{A.37}$$

with $\mathbf{A} \equiv \boldsymbol{a}_\ell \boldsymbol{a}_\ell^T$. The computation of the second derivative yields, after taking the expectation value

$$F_{\alpha\beta}(\ell) = -\left\langle \ln L(\boldsymbol{x}|\boldsymbol{\theta})_{,\alpha\beta} \right\rangle = \frac{1}{2} \operatorname{Tr}\left(\mathbf{S}^{-1}\mathbf{S}_{,\alpha}\mathbf{S}^{-1}\mathbf{S}_{,\beta}\right) , \tag{A.38}$$

where $\mathbf{S} = \langle \mathbf{A} \rangle$ and I have used the following matrix identities:

$$(\ln|\mathbf{S}|)_{,\alpha} = \operatorname{Tr}(\mathbf{S}^{-1}\mathbf{S}_{,\alpha}), \qquad (\mathbf{S}^{-1})_{,\alpha} = -\mathbf{S}^{-1}\mathbf{S}_{,\alpha}\mathbf{S}^{-1} . \tag{A.39}$$

Now let us specify how the $a_{\ell m}^{Ai}$ have been collected into the vector $\boldsymbol{a}_\ell$. Working at fixed multipole ℓ, the vector $\boldsymbol{a}_\ell$ has a multi-index $I = (m, i, A)$. Each of these indices runs on a different range, and the range of the tomographic index i may depend on the probe A considered. Summarising:

- m varies from $-\ell$ to ℓ;
- i varies from 1 to N_A;
- A varies from 1 to $\mathcal{N}$.

Here $\mathcal{N}$ being the number of probes and N_A the number of tomographic bins for probe A. I choose to order the array $\boldsymbol{a}_\ell$ varying the three indices with a significance increasing from left to right. With this choice, the matrix $\mathbf{S}$ has the following block form:

$$\mathbf{S} = \boldsymbol{\Sigma}(\ell) \otimes \mathbb{1}_{2\ell+1}, \qquad \boldsymbol{\Sigma}(\ell) = \begin{pmatrix} \boldsymbol{\Sigma}^{11}(\ell) & \boldsymbol{\Sigma}^{12}(\ell) & \cdots & \boldsymbol{\Sigma}^{1\mathcal{N}}(\ell) \\ \boldsymbol{\Sigma}^{21}(\ell) & \boldsymbol{\Sigma}^{22}(\ell) & \cdots & \boldsymbol{\Sigma}^{2\mathcal{N}}(\ell) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}^{\mathcal{N}1}(\ell) & \boldsymbol{\Sigma}^{\mathcal{N}2}(\ell) & \cdots & \boldsymbol{\Sigma}^{\mathcal{N}\mathcal{N}}(\ell) \end{pmatrix} . \tag{A.40}$$

Using the properties of the Kronecker product it follows that

$$\begin{aligned} \operatorname{Tr}\left[\mathbf{S}^{-1}\mathbf{S}_{,\alpha}\mathbf{S}^{-1}\mathbf{S}_{,\beta}\right] &= \operatorname{Tr}\left[\left(\boldsymbol{\Sigma}(\ell)^{-1}\boldsymbol{\Sigma}(\ell)_{,\alpha}\boldsymbol{\Sigma}(\ell)^{-1}\boldsymbol{\Sigma}(\ell)_{,\beta}\right) \otimes \mathbb{1}_{2\ell+1}\right] \\ &= \operatorname{Tr}\left(\boldsymbol{\Sigma}(\ell)^{-1}\boldsymbol{\Sigma}(\ell)_{,\alpha}\boldsymbol{\Sigma}(\ell)^{-1}\boldsymbol{\Sigma}(\ell)_{,\beta}\right) \operatorname{Tr}\left(\mathbb{1}_{2\ell+1}\right) \\ &= (2\ell+1) \operatorname{Tr}\left(\boldsymbol{\Sigma}(\ell)^{-1}\boldsymbol{\Sigma}(\ell)_{,\alpha}\boldsymbol{\Sigma}(\ell)^{-1}\boldsymbol{\Sigma}(\ell)_{,\beta}\right) \\ &= (2\ell+1) \operatorname{Tr}\left(\boldsymbol{\Sigma}(\ell)^{-1}\mathbf{C}(\ell)_{,\alpha}\boldsymbol{\Sigma}(\ell)^{-1}\mathbf{C}(\ell)_{,\beta}\right) . \end{aligned} \tag{A.41}$$

In the last line I have used the fact that, in this model, the shot noise does not depend on the cosmological parameters. The $C(\ell)$'s block matrix $\mathbf{C}(\ell)$ has the same form of $\boldsymbol{\Sigma}(\ell)$ of Eq. (A.40), with entries $C_{ij}^{AB}(\ell)$ instead of $\Sigma(\ell)_{ij}^{AB}$. From Eq. (A.38) I can finally write

$$F_{\alpha\beta}(\ell) = \frac{2\ell+1}{2} \operatorname{Tr}\left(\boldsymbol{\Sigma}(\ell)^{-1}\mathbf{C}(\ell)_{,\alpha}\boldsymbol{\Sigma}(\ell)^{-1}\mathbf{C}(\ell)_{,\beta}\right) . \tag{A.42}$$

Now, I sum over the multipoles ℓ, considering them as independent:

$$F_{\alpha\beta} = \sum_\ell \frac{2\ell+1}{2} \operatorname{Tr}\left[\boldsymbol{\Sigma}(\ell)^{-1}\mathbf{C}(\ell)_{,\alpha}\boldsymbol{\Sigma}(\ell)^{-1}\mathbf{C}(\ell)_{,\beta}\right] , \tag{A.43}$$

thus obtaining the generic form for the second order covariance formula.

A.4.2 The estimator perspective

In the previous section I have evaluated the expression of the Fisher matrix for a multivariate normal distribution; then, I specialised this expression for the field perspective, when the observables are the $a_{\ell m}$'s. In this section I will derive the Fisher matrix expression when the observables are the estimator $\hat{C}(\ell)$'s, defined as

$$\hat{C}_{ij}^{AB}(\ell) \equiv \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} a_{\ell m}^{Ai} a_{\ell m}^{Bj} \,. \tag{A.44}$$

One may argue that, since the $a_{\ell m}$'s are normally distributed, also the $\hat{C}(\ell)$ should be normally distributed. However, this is not true; in the following I obtain the $\hat{C}(\ell)$ probability density function (pdf) together with its corresponding Fisher matrix. Since $C(\ell)$'s with different ℓ are assumed to be uncorrelated, I will work at fixed ℓ. The $\hat{C}(\ell)$'s likelihood is

$$L(\hat{\boldsymbol{C}}|\boldsymbol{\theta}) = \int d\boldsymbol{a}_\ell L(\boldsymbol{a}_\ell|\boldsymbol{\theta}) \prod_{A,B} \prod_{i,j} \delta \left(\hat{C}(\ell)_{ij}^{AB} - \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} a_{\ell m}^{Ai} a_{\ell m}^{Bj} \right) . \tag{A.45}$$

Using the Dirac delta function the $L(\boldsymbol{a}_\ell|\boldsymbol{\theta})$ can be taken out of the integral

$$L(\hat{\boldsymbol{C}}|\boldsymbol{\theta}) = \frac{f(\hat{\boldsymbol{C}})}{\sqrt{(2\pi)^n |\mathbf{S}|}} \exp\left(-\frac{1}{2} \operatorname{Tr}(\mathbf{S}^{-1}\mathbf{A}) \right) , \tag{A.46}$$

where $f(\hat{\boldsymbol{C}})$ is the result of the integral in the a_{lm}'s space. Taking the logarithm of the previous equation it follows

$$\ln L(\hat{\boldsymbol{C}}|\boldsymbol{\theta}) = \ln\left(\frac{f(\hat{\boldsymbol{C}})}{\sqrt{(2\pi)^n}} \right) - \frac{1}{2} \ln(|\mathbf{S}|) - \frac{1}{2} \operatorname{Tr}(\mathbf{S}^{-1}\mathbf{A}) \,. \tag{A.47}$$

Proceeding as in the previous section and using $\langle \mathbf{A} \rangle = \mathbf{S}$ the expectation value of the second derivative becomes

$$\left\langle \ln L(\hat{\boldsymbol{C}}|\boldsymbol{\theta})_{,\alpha\beta} \right\rangle = -\frac{1}{2} \operatorname{Tr}(\mathbf{S}^{-1}\mathbf{S}_{,\beta}\mathbf{S}^{-1}\mathbf{S}_{,\alpha}) \,. \tag{A.48}$$

Finally, performing the same manipulations of the previous section Eq. (A.43) follows. This proves the equivalence between the field and the estimator perspectives.

Appendix B

SEYFERT: SurvEY FishEr foRecast Tool

The forecast that has been presented in Chapter 4 has been performed using a code named `SEYFERT`. The code is written in modern Python3 – it consists of 11 414 lines of Python code, excluding blank lines and comments[1] – and it has been developed by Luca Paganin (me) and Marco Bonici, under the scientific supervision of Melita Carbone. The structural design of the code, especially at the beginning of the work, was supervised by Stefano Davini. The code is able to compute the Fisher matrices for the *Euclid* main probes, i.e. WL, $\mathrm{GC_{ph}}$, and $\mathrm{GC_{sp}}$, together with all their possible cross-correlations. The observables employed are the angular power spectra in the Limber approximation, while the computation of the Fisher matrix is performed using the recipe detailed in Sec. 4.2. The code is also able to take an external Fisher matrix as input, as it has been done to implement the so-called hybrid approach explained in Sec. 4.2.3. In this case the $\mathrm{GC_{sp}}(P_k)$ Fisher matrix computed by the IST Forecast [5] from the spectroscopic galaxy-galaxy Fourier power spectrum has been taken as an external input.

B.1 Main concepts and classes

In this section the main concepts that have been implemented in the code are described. Each of these concepts has been implemented in a `SEYFERT` Python class, and belongs to a specific sub-module of the `SEYFERT` Python package, which are briefly described in the following.

B.1.1 Core terms

Probe With the word "probe" it is denoted a generic cosmological entity that can be associated to a field f that can be expanded in spherical harmonics, like in Eq. (4.1). In this sense, the probes considered in the work presented in Chapter 4 are 3: weak lensing (WL), photometric galaxy clustering ($\mathrm{GC_{ph}}$), and spectroscopic galaxy clustering ($\mathrm{GC_{sp}}$). Capital letters in alphabetic order (A, B, C,...) are used to indicate a generic probe. In `SEYFERT` a probe is identified by mean of a single string, `probe_key`, which simply stores the name of the probe itself.

[1] To count the lines of code I used the open-source tool `cloc` `https://github.com/AlDanial/cloc`.

Observable The word "observable" is used to denote a generic two-point function given by the correlation of two probes. Specifically, in `SEYFERT` the observables are the angular power spectra, i.e. the harmonic transform of the two-point correlation function.

Data-vector A "data-vector" is a list of the observables used to compute a Fisher matrix. If a data-vector contains more than one observable, all the pairwise cross-covariances between the included observables are included in the covariance matrix associated to the data-vector itself.

B.1.2 Cosmology module

This `SEYFERT` module collects the class implementations for the computation of all the relevant cosmological quantities. The module contains the source code for the execution of the first two `SEYFERT` tasks, i.e. `PowerSpectrum` and `Angular` (see Sec. B.2).

`Cosmology` This is one of the most important classes of the Cosmology module. It contains the values of the cosmological parameters, and implements methods to compute the Hubble parameter $H(z)$, the comoving distance $\chi(z)$, and the growth factor $D(z)$. It also possesses an instance of the `PowerSpectrum` class and can be invoked to compute the power spectra. This class is possessed by all the other classes representing derived quantities which depend on the above mentioned basic cosmological functions.

`PowerSpectrum` This class implements the concept of matter power spectrum. It is used for storing linear and nonlinear matter power spectra, and it can also be invoked to compute them. The computation is done via a call to the selected Boltzmann solver, which can be either `CAMB` or `CLASS`. The power spectra are stored as 2D `numpy` arrays, with the two axes representing the redshift and the wave-number dimensions respectively. This object also stores the redshift and the wave-number grids on which the power spectra are evaluated by the Boltzmann solver. The most important pieces of data that this class manages are the following:

- `z_grid`: a 1D `numpy` array representing the redshift grid on which the power spectrum is evaluated.
- `k_grid`: a 1D `numpy` array representing the wave-number grid on which the power spectrum is evaluated.
- `lin_p_mm_z_k`: a 2D `numpy` array representing the *linear* matter power spectrum evaluated on the 2D grid defined by `z_grid` and `k_grid`.
- `nonlin_p_mm_z_k`: a 2D `numpy` array representing the *non-linear* matter power spectrum evaluated on the 2D grid defined by `z_grid` and `k_grid`.

`PhysicalParameter` This is a basic simple class implementing the concept of a physical parameter of the forecast, meaning that it can be a cosmological parameter or a nuisance parameter. A parameter can be free to vary, in the sense that derivatives with respect to it can be computed, or it can be fixed.

`RedshiftDensity` This class represents the redshift distribution of a given cosmological probe, and as such is uniquely labeled by a `probe_key`. More concretely, the resdshift density is associated to a particular tracer of the dark matter distribution. The most important task of this class is to provide the tomographic normalised density distribution $n_i^A(z)$ defined in Eq. (4.5). In order to do the task, it implements an instrument response parameterised by the double gaussian defined in Eq. (4.6). The main data structure contained in this class is a 2D `numpy` array `n_iz` storing the values of the tomographic normalised density $n_i^A(z)$ evaluated on the `PowerSpectrum` redshift grid.

`WeightFunction` This class implements the concept of the tomographic weight function $W_i^A(z)$ appearing in the Limber integral Eq. (4.4). Three subclasses of this class exists, implementing the $\mathrm{GC_{ph}}, \mathrm{GC_{sp}}$, and WL weight functions defined by equations Eq. (4.12), Eq. (4.14), and Eq. (4.21) respectively. A `WeightFunction` is uniquely identified by its probe index A (the `probe_key`). The class possesses an instance of the `Cosmology` class and an instance of the `RedshiftDensity` class. The `probe_key` of the density must be equal to the one of the weight function, with the only exception of the `LensingWeightFunction`, i.e. the weight function associated to WL Eq. (4.21). This is because the WL is computed by using the redshift density of $\mathrm{GC_{ph}}$.

The most important piece of information carried by the `WeightFunction` class is a 2D `numpy` array `w_iz` with the values of $W_i^A(z)$ evaluated on the `PowerSpectrum` redshift grid.

`KernelFunction` This is mainly a convenience class which defines the kernel function against which the matter power spectrum is integrated in Eq. (4.4). The kernel functions are defined as in [5], i.e.

$$K_{ij}^{AB}(z) \equiv c\frac{W_i^A(z)W_j^B(z)}{H(z)\chi^2(z)} \tag{B.1}$$

The main task of the `KernelFunction` class is to compute and store a 3D `numpy` array `k_ijz` containing the values of the kernel. This array is then multiplied by the 2D array containing the values of the non-linear matter power spectrum and it is integrated along the redshift dimension.

`AngularCoefficient` This class represents a tomographic angular power spectrum $C_{ij}^{AB}(\ell)$. As such it possesses a `KernelFunction` instance, which in turn comes with a `Cosmology` instance and two `WeightFunction` instances. An `AngularCoefficient` is identified by mean of two strings, which are the strings identifying the two weight functions, i.e. two `probe_key`'s
.

The main operations performed by the `AngularCoefficient` class are the following:

- evaluate the power spectrum in Limber approximation on the given multipoles ℓ grid. The computation is performed via a two-dimensional interpolation of the non-linear matter power spectrum, which is defined on a z-k grid. The result is stored in the attribute `limber_power_spectrum_l_z`, and it is a 2D array: the two dimensions being the multipole and redshift dimensions respectively.

- evaluate the kernel function using the already evaluated weight functions, which are computed and stored before starting the computation of the $C(\ell)$'s themselves (see `AngularCoefficientsCollector` paragraph).
- multiply the kernel and the Limber power spectrum, and integrate over the redshift dimension using numerical Newton-Cotes discrete methods (see Sec. B.4). This method is way faster than interpolation and then integration via Gauss quadrature rules (see discussion in Sec. B.4).

The result of the computation is stored in a 3D `numpy` array `c_lij`, which stores the values of the power spectra for each combination ℓ, i, j. The name of the attribute is self-descriptive about the ordering of the axes, i.e. ℓ, i, and j are respectively the first, second, and third dimension of the array.

`AngularCoefficientsCollector` This a collector class whose aim is to orchestrate the computation of all angular power spectra that can be constructed with the selected cosmological probes. The weight functions of the probes are evaluated once and then are re-used for computing the kernel functions of the angular power spectra. This is the main class that is instantiated and invoked to execute the `Angular` task. The main steps carried out by this class are

- Load the cosmology evaluated by the `PowerSpectrum` task, with the already computed matter power spectra.
- Evaluation and storage of the weight functions for the present probes.
- Loop over all the possible angular power spectra, computing them efficiently by re-using the already computed weight functions and matter power spectra.
- Store the results on disk in an HDF5 file.

B.1.3 Derivatives module

This module is responsible for the computation of the $C(\ell)$'s derivatives with respect to the free parameters of the forecast. The computation is done via the SteM fitting procedure described in Sec. B.5 [82]. The computation of the derivative with respect to a parameter relies on the already computed spectra for 15 different values of the parameter itself, which are the output of the `Angular` task.

`SteMClDifferentiator` The class is responsible for the computation of the derivative itself. It loads the angular power spectra for all the values of the current parameter, and for each combination of ℓ, i, j it performs the SteM iterative linear fit, yielding the slope of the regression as the value of the derivative. In practice the input for the computation of the derivative is a 4-dimensional array `c_dvar_lij`, which represents a stacked collection of `c_lij` arrays along the first dimension, which corresponds to the physical parameter (`dvar`) in question. The linear regression must be performed along the first dimension of the array, for all values of the last three indices. Thus it is not straightforward to implement this procedure in the most efficient way. The algorithimc details of the implementation that has been done in `SEYFERT` are reported in Sec. B.5.

`ClDerivative` This class acts as container for the values of the derivative $C_{ij}^{AB}(\ell)_{,\theta}$ of a single angular power spectrum with respect to a free parameter θ. It possesses an instance of `SteMClDifferentiator`, which is used to evaluate the derivative, and stores the result of the computation in a 3D `numpy` array `dc_lij`. The ordering of the axes is the same of the `c_lij` array stored by the `AngularCoefficient` class.

`ClDerivativeCollector` This class is a collector of `ClDerivative` instances, and it is the main entry point of the `Derivative` task. The derivatives are evaluated through a call to a method of this class, which then stores them in a python dictionary. The results are then saved to disk on a HDF5 file.

B.1.4 Fisher module

This is the module devoted to the computation of the Fisher matrix elements, together with the final results of the forecast obtained from the inversion of the Fisher matrices themselves.

`ClDataVector` This is a simple class defining a the structure of the data-vector for the computation of a given Fisher matrix. It is essentially a list of pairs (A, B) of `probe_key`'s, each one identifying a kind of angular power spectrum. In symbols it can be written as

$$[(\mathrm{A}_1, \mathrm{B}_1), (\mathrm{A}_2, \mathrm{B}_2), \ldots, (\mathrm{A}_\mathrm{N}, \mathrm{B}_\mathrm{N})] \,, \tag{B.2}$$

where N is the number of pairs involved. The most important task of this class is to establish the block ordering structure of the covariance matrix of the $C(\ell)$'s, along with the ordering of the vectorised derivatives of the angular power spectra to be used in the computation of the Fisher matrix elements. Each `ClDataVector` has a name, that can be deduced from its entries according to the rules defined in Sec. 4.2.1. In practice, the rules for building the name are the following:

1. the pair (A, A) is mapped to the string "A";

2. the pair (A, B) with $\mathrm{A} \neq \mathrm{B}$ is mapped to the string "XC(A, B)";

3. the full name is built by mapping all pairs into strings according to the two rules above and concatenating all the strings with plus signs. The final result is wrapped with square brackets, in order to delimit the datavector extent.

Some examples:

$$\begin{aligned}
[(\mathrm{WL}, \mathrm{WL})] &\longrightarrow \text{“}[\mathrm{WL}]\text{”} \\
[(\mathrm{WL}, \mathrm{WL}), (\mathrm{GC_{ph}}, \mathrm{GC_{ph}})] &\longrightarrow \text{“}[\mathrm{WL} + \mathrm{GC_{ph}}]\text{”} \\
[(\mathrm{WL}, \mathrm{WL}), (\mathrm{WL}, \mathrm{GC_{ph}})] &\longrightarrow \text{“}[\mathrm{WL} + \mathrm{XC}(\mathrm{WL}, \mathrm{GC_{ph}})]\text{”} \\
[(\mathrm{WL}, \mathrm{WL}), (\mathrm{GC_{ph}}, \mathrm{GC_{ph}}), (\mathrm{WL}, \mathrm{GC_{ph}})] &\longrightarrow \text{“}[\mathrm{WL} + \mathrm{GC_{ph}} + \mathrm{XC}(\mathrm{WL}, \mathrm{GC_{ph}})]\text{”} \,.
\end{aligned} \tag{B.3}$$

`ClCovarianceMatrix` This class represents the covariance matrix of the $C(\ell)$'s. This is implemented through a 3D `numpy` array: the first dimension represent the multipoles in which the $C(\ell)$'s are sampled. The last two dimensions are two multi-indices representing the vectorisation of the tomographic indices pairs of the involved $C(\ell)$'s. More explicitly, for a fixed ℓ and a given pair A, B the $C_{ij}^{AB}(\ell)$ is a tomographic matrix. This matrix is symmetric if $A = B$, while when $A \neq B$ it is not even necessarily square as well as not symmetric. To build the covariance matrix it is necessary to turn this matrix into a vector, using the vec or vecp operators introduced in Sec. A.2. This is realised through a mapping between a pair of tomographic indices i, j and a single multi-index I. More specifically, assuming $A \neq B$:

$$[\mathrm{vecp}(C^{AA}(\ell))]_I = C_{ij}^{AA}(\ell), \qquad [\mathrm{vec}(C^{AB}(\ell))]_I = C_{ij}^{AB}(\ell)\,. \tag{B.4}$$

The range of the multi-index I can be denoted as N_{AB}, and

$$N_{AA} = \frac{N_A(N_A+1)}{2}, \qquad N_{AB} = N_A \cdot N_B\,. \tag{B.5}$$

With this notation, the block AB-CD of the covariance matrix is a $N_{AB} \times N_{CD}$ matrix, and can be rewritten as

$$\left\{\mathrm{Cov}\left[C^{AB}(\ell), C^{CD}(\ell)\right]\right\}_{IJ} = \mathrm{Cov}\left[C_{ij}^{AB}(\ell), C_{km}^{CD}(\ell)\right]\,. \tag{B.6}$$

Each of these blocks is used to build the full covariance matrix, with a layout which is based on the data-vector ordering, as the examples in Sec. 4.2.1 show. In principle this block structure is equivalent to the definition of a quadruple multi-index, mapping (A, B, i, j) to a single number. Both solutions, if correctly implemented, should give the same answer.

`FisherEvaluator` This is a class whose task is to compute the Fisher matrix elements for a given data-vector. It takes the necessary inputs for building the covariance matrix and the $C(\ell)$'s derivative vectors, and performs the computation of the Fisher matrix using Eq. (4.34) and Eq. (4.35).

`FisherMatrix` This class represents the Fisher matrix. The matrix is stored in the `matrix` attribute, which is a `pandas DataFrame`, a data structure allowing to label rows and columns of a table using strings. This class also stores the name of the Fisher matrix, together with two python `set`s specifying which of the parameters are nuisance and which are cosmological.

`FisherAnalysis` This class works as a container, holding a dictionary of `FisherMatrix` objects. It also implements methods for computing the parameters of the gaussians and the ellipses that are drawn in the contour plots.

B.2 Dataflow and execution modes

In this section it is described the dataflow of the code, along with the execution modes available.

The dataflow of `SEYFERT` can be subdivided in four main different computational *tasks*:

1. `PowerSpectrum`: computation of matter power spectra;

2. `Angular`: computation of angular power spectra;
3. `Derivative`: computation of the derivatives of the angular power spectra;
4. `Fisher`: computation of the Fisher matrices.

These tasks are described in detail in the next subsection. A complete run of the code comprises the execution of all the above described tasks, which must be run in the correct order. However, it is not necessary to run all the steps at once. In fact, one of the main strengths of `SEYFERT` is its high *modularity*, i.e. the possibility to run each task separately, taking the necessary inputs from the output of the previous steps. In particular, for the forecast described in Chapter 4, the power spectra were not computed with `SEYFERT`. Instead the IST:F power spectra [5] have been taken as external inputs, after the appropriate file formats conversions.

B.2.1 Computation tasks

In this section a schematic description of the sequential steps done in each task is provided.

`PowerSpectrum`

This is the step in which the linear and non-linear matter power spectra are evaluated. The computation is done via calling an external Boltzmann solver – choices are among `CAMB` or `CLASS`– and it is repeated for different values of the cosmological parameters. This is because the numerical differentiation algorithm employed for computing the derivatives needs that the $C(\ell)$'s – and thus also the matter power spectra – are computed for different values of each parameter. The computation of the matter power spectra is performed invoking a specific method of the `Cosmology` class described in Sec. B.1.

`Angular`

This is the step in which the angular power spectra are evaluated starting from the matter power spectra. The input power spectra can be either computed by `SEYFERT` in the `PowerSpectrum` task, or taken as external inputs. The computation of the redshift integral is performed via numerical integration with composite Newton-Cotes quadrature rules, described in detail in Sec. B.4. Like the `PowerSpectrum` task, also the `Angular` task must be executed multiple times, separately varying each of the parameters. The computation of the $C(\ell)$'s is performed invoking a specific method of the `AngularCoefficientsCollector` class described in Sec. B.1, and it is repeated for all kind of $C(\ell)$'s that can be constructed from the probes present in the forecast.

`Derivative`

Here the derivatives of the angular power spectra are computed using the SteM numerical differentiation method, which is detailed in Sec. B.5. The varied angular power spectra are loaded from the output of the `Angular` task, and then the SteM fitting procedure is applied for each parameter. The derivatives are computed both for the cosmological and for the nuisance parameters. The class instantiated by this task is `ClDerivativeCollector`, with which then the derivatives with respect to a given parameter are computed. This is done for all kind of $C(\ell)$'s that can be constructed from the probes present in the forecast.

`Fisher`

Here the Fisher matrices are computed, using the input of the previous `Angular` and `Derivative` steps. The task accepts more than one data-vector in the form of a instance of the `ClDataVector` class. The computation is done using the `FisherEvaluator` class. For a given data-vector, this class builds the covariance matrix, along with the vectorised derivatives, and then applies the formula Eq. (4.35). A sum over the multipoles is done to obtain the final Fisher matrix elements. The result of the computation are the full Fisher matrices for each of the data-vector that have been specified.

B.2.2 Execution modes

The main entry point of the code is the `job_submitter` script, which among the other options accepts the execution mode through the `--execution` command line argument. The code can be executed in two ways: `interactive` and `batch`.

Interactive execution

The `interactive` way can be run on any machine, and it is fully sequential. This especially affects the performances of the `PowerSpectrum` and `Angular` tasks. These tasks must be repeated multiple times for each free parameter, in order to produce the variations for the computation of the derivatives of the angular power spectra. Since in the interactive mode all these calculations are done sequentially, this mode is slower than the batch execution, with which multiple jobs are executed in parallel. This mode is intended to be used for debugging purposes, since it allows a step-by-step monitoring of the code execution, or if the user does not have access to an LSF cluster. The runtime of this mode varies significantly depending on the configuration chosen. In particular it takes about 22 minutes when using 4 bins for GC_{sp}, which is the lightest configuration. Instead it takes 83 minutes in the most time-consuming settings, i.e. when using 40 GC_{sp} bins. The runtimes reported here for the interactive mode have been obtained by running on a workstation with Ubuntu 20.04 LTS, with a RAM of 32 GB and a Intel Core CPU i7-4820K with a clock frequency of 3.70 GHz.

Batch execution

The `batch` mode allows to parallelise the independent computational steps, but needs to be run on a cluster equipped with the LSF orchestration system [90]. The parallelism is done via submitting multiple jobs with the `bsub` LSF command. The parallel jobs are the ones belonging to the same task, while the four tasks must be run one after another. This is because each task takes as input the outputs of one or more of the previous tasks. The batch execution is more efficient than the interactive, taking about 10 minutes to run with 4 GC_{sp} bins, and ~ 40 minutes when the bins are 40. This means the batch execution is faster than the interactive roughly by a factor of ~ 2. The runtimes reported here for the batch mode have been obtained by running on the INFN Genova computing facility.

Figure B.1: Example view of the forecast configuration `JSON` file.

B.3 File formats

This section briefly describes the file formats adopted in `SEYFERT`, both for the configuration and the output data storage.

B.3.1 Configuration files

All `SEYFERT` configuration files are written in JavaScript Object Notation (`JSON`) format. This format has been chosen for the following reasons:

- `JSON` files are essentially collections of key-value pairs, therefore each entry must have a name and a value. This comes automatically with readability, since the keys are self-descriptive of what they point to (the values). Moreover `JSON` supports all the basic types like `float`, `string` and `bool`.

- Python has a standard library (`json`) that is very easy to use, and allows to do `JSON` file I/O operations with few lines of code. A `JSON` file can be easily read into a Python dictionary, which is a built-in data structure. Viceversa, a python dictionary can be easily dumped to a `JSON` file.

- `JSON` can be *nested*. That is, the main dictionary can contain other dictionaries as values of its keys. This allows to put highly structured information into a single file.

- `JSON` is a simple text format that can be easily read, written, and maintained. If a new option or a new group of options is needed it is sufficient to add a new entry with a new value, which can also be a dictionary of values. This operation won't affect any of the other options that were already present.

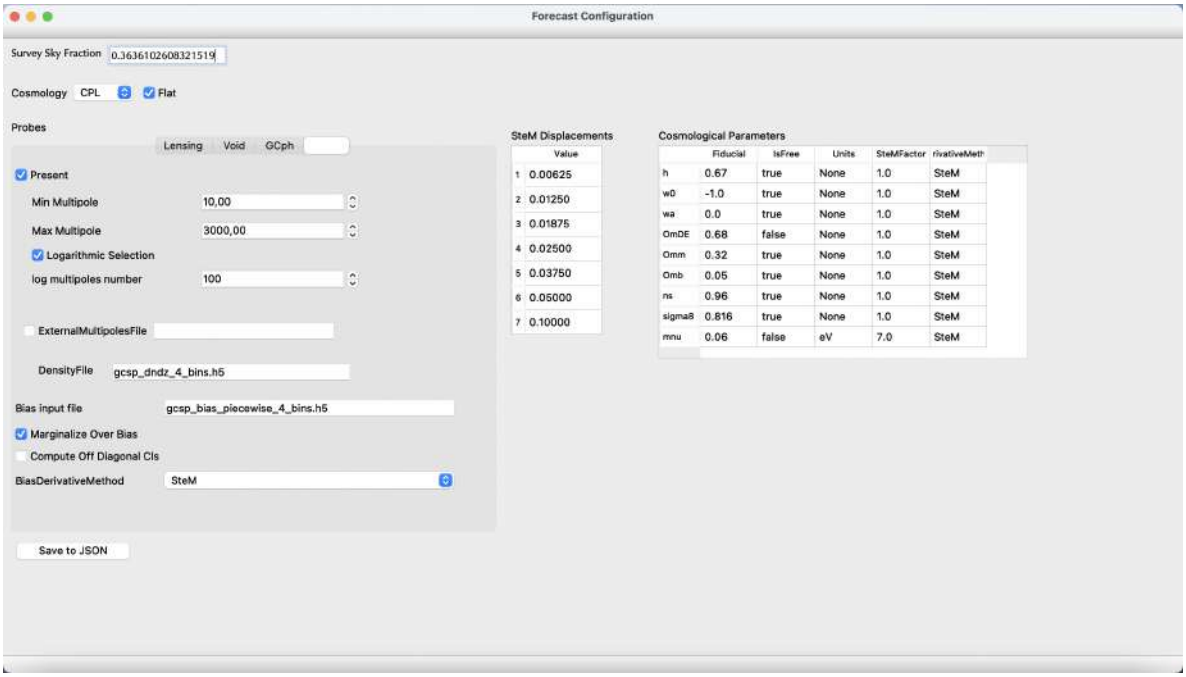

Figure B.2: Example view of the forecast configuration GUI.

In `SEYFERT` there are two kinds of configurations files: the forecast configuration and the task configurations.

Forecast Configuration

The forecast configuration is a single `JSON` file, which specifies the global configuration of the forecast. In Fig. B.1 it is shown an example view of the forecast configuration. The file is highly structured, being 240 lines long, so it cannot be expanded on a single screen view. The main keys are shown in the figure:

- `metadata`: metadata describing the file itself, like author, date and time of creation and version;
- `synthetic_opts`: these are shortcut high-level options that are processed to create the configuration file automatically. For example if `scenario` is set to `optimistic`, the code automatically sets the optimistic multipole ranges for the present probes. However, these options are only used as a shortcut for the automatic creation of the file, and are written to the file for descriptive reasons. After the file is created the other settings can be modified manually.
- `survey`: general options about the survey, e.g. the sky fraction covered.
- `cosmology`: settings of the cosmology, i.e. flatness and the settings of the cosmological parameters.
- `probes`: this a nested dictionary containing the settings for the single probes of the forecast.

SEYFERT has also a graphic user interface (GUI) window that allows to create and modify the configuration file in a user-friendly way. An example view of the GUI is shown in Fig. B.2. After editing the settings the user can click on the "Save to JSON" button: this action will trigger the opening of a dialog window, in which the user can insert the name of the file to be saved. After finishing, the configuration will be dumped to a JSON file ready to be used by the code.

Task configurations

Each of the SEYFERT tasks has configuration file, containing options that are specific to the computation that is performed in that task.

PowerSpectrumConfig The configuration for the PowerSpectrum task. It essentially contains high-level options for the Boltzmann code, setting the sampling and the extent of the k and z grids. It also contains the name of the Boltzmann code to use, i.e. CAMB or CLASS.

AngularConfig The configuration for the Angular task. It specifies the order of the integration quadrature rule, together with the interpolation settings for computing the Limber approximated matter power spectrum.

DerivativeConfig The configuration for the Derivative task. In past versions of the code this was used, but now it is empty since there was no need of specific options for this task.

FisherConfig The configuration for the Fisher task. Like the DerivativeConfig it was used in previous versions of SEYFERT, while now it is not being used any more.

B.3.2 Data files

The output data of SEYFERT are almost always stored in binary HDF5 format. The main advantages of this format are the following:

- It can contain natively numpy arrays, which are widely and almost omni-present in SEYFERT.
- It supports various algorithms for data compression, such as gzip or bz2.
- An HDF5 can be arbitrarily nested, resembling the structure of a directory tree with subdirectories (HDF5 groups) and files (HDF5 datasets).
- There exists a Python library (h5py) that can be used for easily creating, reading and writing HDF5 files.
- Multi-dimensional numpy arrays can be natively stored as HDF5 datasets, allowing to save and load them without any intermediate reshape or transformation.
- HDF5 datasets and groups can have a dictionary of *attributes*, which can be used to store simple data as bool, string or float. Attributes are also useful to store the metadata of the files itself, like creation date, description and so on.

- The datasets stored in an HDF5 file are retrieved only when explicitly accessed. This allows to save RAM usage, since opening the file does not automatically load all the file content in memory. This allows to work with very large files, as long as single datasets are small enough to fit in the RAM.

Two kinds of additional input HDF5 files are needed for a `SEYFERT` run, containing information about the redshift density and the bias (if present) of a given probe. Examples of these files can be found in the `SEYFERT` gitlab repository.

B.4 Numerical integration

The integrals that define the angular power spectra in the Limber approximation are evaluated numerically. Since the matter power spectra are evaluated from the Boltzmann solver on a regular redshift grid, an interpolatory quadrature rule must be employed in order to compute the integral. Therefore the problem is to approximate the value of the integral knowing the integrand only at regularly sampled points.

There are various approaches that can be followed, but in this case some kind of interpolation must be done, explicitly or implicitly. The simplest quadrature rules are the Newton-Cotes formulas (see Sec. B.4.1), which give an approximation of the integral with a weighted sum of the function samples. This weighted sum is the closed-form expression of the integral of a interpolating polynomial, so there is an implicit interpolation in this method.

More precise methods for approximating integrals are Gaussian quadrature rules with free nodes [91]. These rules rely on evaluating the integrand function at particular points, i.e. the nodes of the quadrature, which are adjusted to minimise the error for the given function. This means that the positions of these nodes are not a priori known, and most importantly they depend on the particular function in exam. Therefore an interpolation must be done explicitly, and then the interpolating function (spline) has to be passed to the quadrature algorithm, which optimally finds the free nodes. Moreover, the result is an accurate approximation of the *integral of the spline*, which is not necessarily closer to the true (unknown) value than the result given by the Newton-Cotes quadrature. Therefore, in this case the Gauss quadrature is not necessarily better than the classical Newton-Cotes quadrature. Moreover, Newton-Cotes rules give closed-form expressions that can be easily vectorised, and this significantly boosts the performances. For this reason the choice that has been done in `SEYFERT` is to adopt the Newton-Cotes formulas for integrating Eq. (4.4) numerically.

B.4.1 Newton-Cotes quadrature

Newton-Cotes quadratures formulas [91] are interpolatory rules that approximate the integral of a function $f(x)$ using a weighted sum of the function samples on a regular grid. Assume to have sampled f on n regularly spaced points $x_1, \dots, x_n$, with corresponding function values $f_i = f(x_i)$. An interpolatory quadrature formula consists in approximating the integral of f as the integral of a polynomial $p(x)$, which can be written as a weighted sum of the function samples f_i

$$\int_a^b p(x)\mathrm{d}x = \sum_{i=1}^{n} w_i f_i \,, \tag{B.7}$$

where $p(x)$ is the unique polynomial of degree $n-1$ interpolating $f(x)$ at the distinct points $x_1, x_2, \ldots, x_n$. By Lagrange's interpolation formula

$$p(x) = \sum_{i=1}^{n} f(x_i)\ell_i(x), \quad \ell_i(x) = \prod_{\substack{j=1 \\ j\neq i}}^{n} \frac{(x - x_j)}{(x_i - x_j)} \tag{B.8}$$

where $\ell_i(x)$ is the elementary Lagrange polynomials associated with the node x_i. It follows that the weights are given by

$$w_i = \int_a^b \ell_i(x)\mathrm{d}x \tag{B.9}$$

Newton-Cotes formulas are special cases of interpolatory rules when the nodes are equispaced.

In general, having n nodes uniquely determines a polynomial of degree $n-1$. However it is known that Newton-Cotes formulas becomes numerically unstable very soon with increasing the order of the interpolating polynomial. In particular this starts to happen already with $n \geq 8$, see [92] page 390 for a discussion. Of course, in real life more than 8 function samples will be available. In practice what is done is to use *composite* low order rules, which consist in applying the interpolation to small sub-intervals of the integration range, and then sum up together all the contributions to get a global answer for the integral.

In `SEYFERT` by default a composite Newton-Cotes rule of order 2 (Simpson's rule) is used; the maximum order allowed is $n = 6$. In what follows the simplest rules are briefly reviewed, showing the performances on a non-trivial test function when the number of samples is increased.

Trapezoidal rule

After the Riemann sum, which approximates the integral of a function with a sum of rectangles, there is the trapezoidal rule, which is the Newton-Cotes formula of order 1. Given two points $x_1 = a$ and $x_2 = b$ with the corresponding function values f_a, f_b, the interpolating polynomial is a straight line

$$p(x) = f_a + \frac{f_b - f_a}{b - a}(x - a) \tag{B.10}$$

The approximation to the integral in this case is given by the area under this straight line, which is a trapezium

$$\int_a^b \mathrm{d}x\, p(x) = \frac{b - a}{2}(f_a + f_b) \tag{B.11}$$

However this is a very poor approximation of the integral, as the upper left panel of Fig. B.3 shows. When more samples are available, a composite rule can be adopted. This consists in applying the linear interpolation to more than one couple of points. Supposing to have a total of n points including the extrema a and b, there will be $n-2$ internal points x_i

$$x_1 = a, \quad x_i = x_a + (i-1)h, \quad x_n = b, \qquad i = 2, \ldots, n-2\,, \tag{B.12}$$

and the spacing between the points will be $h = \frac{b-a}{n-1}$, since there are $n-1$ sub-intervals between a and b. The trapezoidal rule applied to a single sub-interval $[x_i, x_{i+1}]$ gives a contribution of

$$T_i = \int_{x_i}^{x_{i+1}} \mathrm{d}x\, p(x) = \frac{h}{2}(f_i + f_{i+1}) \tag{B.13}$$

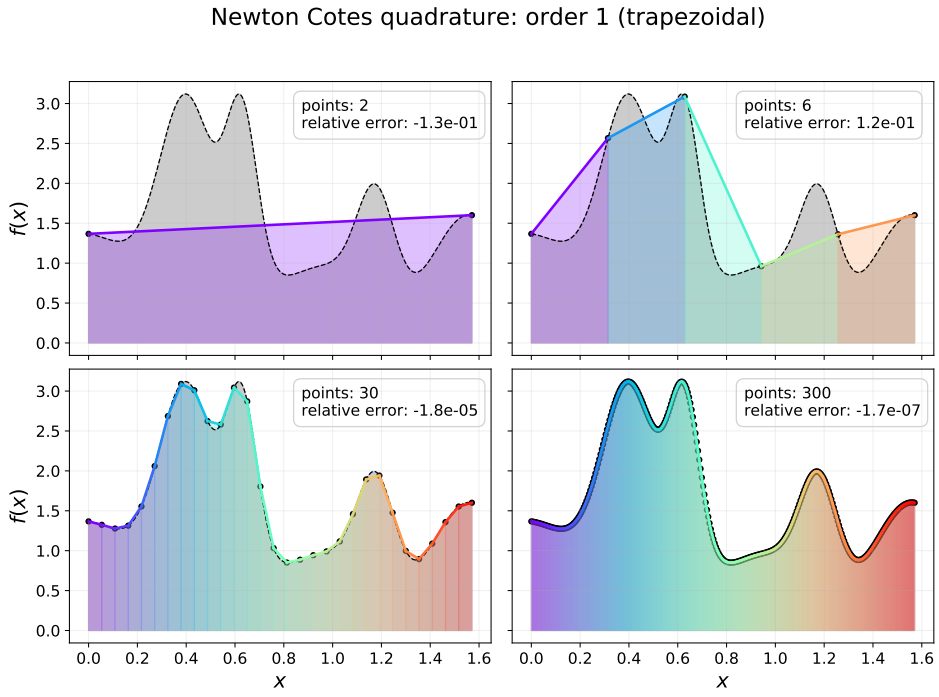

Figure B.3: Composite trapezoidal rule performances with increasing number of samples, applied to a test function.

The total integral can be written as a sum of these pieces coming from the $n-1$ sub-intervals:

$$\int_a^b \mathrm{d}x\, f(x) \simeq \sum_{i=1}^{n-1} T_i = \frac{h}{2}\left[f_1 + 2\sum_{i=2}^{n-2} f_i + f_n\right], \tag{B.14}$$

where the 2 in front of the internal points comes from the fact that these will be shared between adjacent sub-intervals. Instead, this clearly not happen for the extrema, that appears only once in the sum. It can be shown [91] that the global truncation error of the composite trapezoidal rule is proportional to h^2. This means that increasing the number of points gives progressively better results, as it can be intuitively expected. This is shown in Fig. B.3, where the error relative to the exact analytical value of the test function decreases down to 1×10^{-7} when using 300 samples in the interval $[a, b]$.

Simpson's rule

The Newton-Cotes rule of order 2 is also referred as the Simpson's rule, and it is based on a parabolic interpolation of the function samples. For this reason at least 3 samples are needed in order to apply this rule. Supposing to have 3 equispaced points x_1, x_2, x_3, with $x_2 = x_1 + h$ and $x_3 = x_1 + 2h$, the three Lagrange basis polynomials defined by Eq. (B.8) are

$$\ell_1(x) = \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)}, \quad \ell_2(x) = \frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)}, \quad \ell_3(x) = \frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)}\,. \tag{B.15}$$

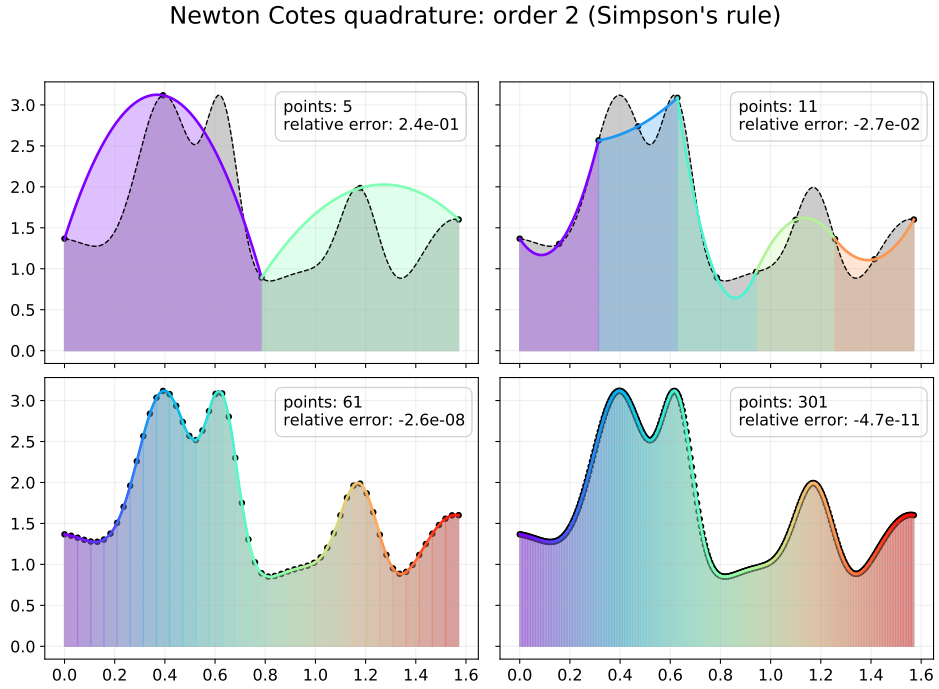

Figure B.4: Composite Simpson's rule applied to a test function, progressively increasing the number of samples.

The interpolating parabola is obtained from a linear combination of these polynomials weighted by the function's values:

$$p(x) = f_1\ell_1(x) + f_2\ell_2(x) + f_3\ell_3(x)\,. \tag{B.16}$$

A straightforward computation of the integral of the ℓ_i's then leads to the Simpson's rule

$$\int_{x_1}^{x_3} \mathrm{d}x\, f(x) \simeq \int_{x_1}^{x_3} \mathrm{d}x\, p(x) = \frac{h}{3}\left(f_1 + 4f_2 + f_3\right)\,. \tag{B.17}$$

Like for the trapezoidal, a composite rule can be built. However, when applying the composite Simpson's rule the samples must be subdivided in groups of 3 points each. Since the first and the last point of each group are shared by the adjacent groups, except for the extrema, it turns out that the number of samples must be *odd* in order to have an integer number of groups. Assuming for the moment to have an integer number of groups, the composite Simpson's rule reads

$$\int_a^b \mathrm{d}x\, f(x) \simeq \frac{h}{3}\sum_{i=1}^{\frac{n-1}{2}}(f_{2i-1} + 4f_{2i} + f_{2i+1}) \tag{B.18}$$

where by definition $x_1 = a$ and $x_n = b$, and h is again the step size given by $h = \frac{b-a}{n-1}$. If the number n of samples is even, one point will be left out of the computation. It can be included by applying the trapezoidal rule to an additional pair of points, including the remaining point and another one belonging to a group of three. How this pair is chosen is

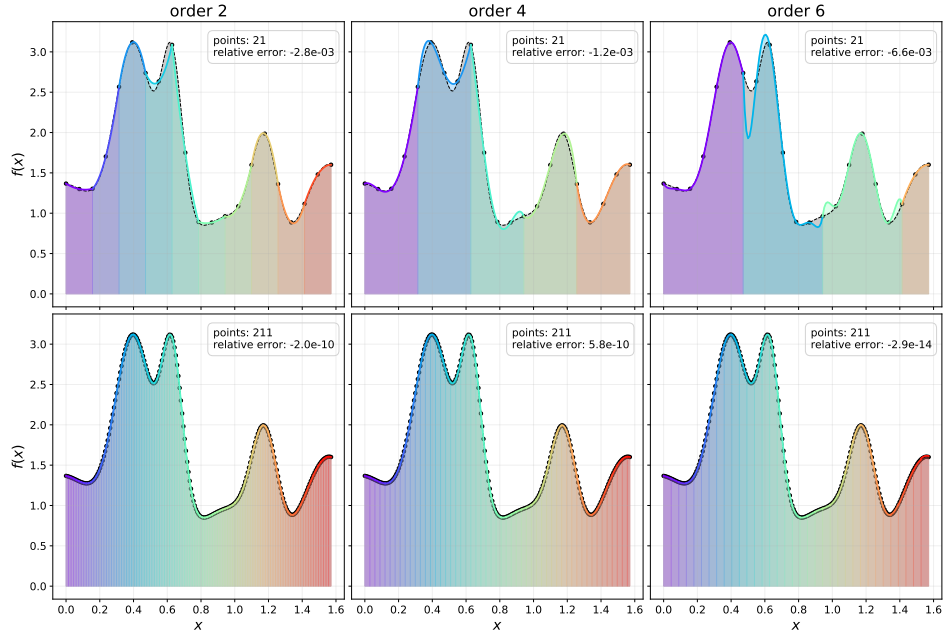

Figure B.5: Comparison of the first three even orders of the Newton-Cotes quadrature. Each column corresponds to a given order, as indicated in the titles.

conventional. The Simpson's composite rule implemented in `scipy` (`scipy.integrate.simps`) does the computation in two ways, applying the trapezoidal rule to the first and second point or to the last and the second-last. The final result is computed by giving a simple average of the two results. The convention followed in `SEYFERT` instead is to apply the trapezoidal rule only to the last and second-last points, since in this way the algorithm can be generalised to higher orders more easily.

The truncation error for the Simpson's rule decreases with the step size with order h^4 [91], which gains two orders of magnitude with respect to the trapezoidal rule, without using more function evaluations.

B.4.2 Implementation of the quadrature algorithm

The Newton-Cotes quadrature has been implemented in `SEYFERT` exploiting again the vectorisation offered by `numpy`. In fact the integrand function of the angular power spectra in general is a 4D array, with the four dimensions corresponding to ℓ, i, j, z respectively. The algorithm has therefore been designed to accept a multi-dimensional array, with the possibility to specify the axis along which the integration has to be performed. Since the approximation of the integral is given by a weighted sum, the vectorisation can be easily achieved by doing a proper indexing of the integrand array. The weights for the given order n are computed once by mean of the `scipy` function `scipy.integrate.newton_cotes`, which gives $n + 1$ weights. For example for $n = 1$ (trapezoidal) the weights will be two and equal to $1/2$, as Eq. (B.13) states. For $n = 2$ (Simpson) the weights are 3 and equal to $1/3, 4/3$, and $1/3$ respectively, and so on. The maximum order allowed is $n = 6$, in order to be safe from the aforementioned high-order numerical instabilities. After the order and the weights are known, the number N_b

of sub-groups is computed from the number of points N_p by an integer division, according to the formula:

$$N_p - 1 = N_b \cdot n + r \,. \tag{B.19}$$

Here r is the integer remainder from the division $(N_p - 1)/n$ and is the number of points that are left out of the composite rule of order n. The first step of the numerical integration is done by applying the composite rule at the first $N_b \cdot n$ points. If $r \neq 0$, Newton-Cotes of order r is applied to the last $r + 1$ points, and the contribution is summed to the previous one. This is because having $r + 1$ samples – the r remaining points and the last of the included ones – allows to apply Newton-Cotes of order r at most. This is because $r + 1$ points allows to uniquely determine an interpolating polynomial with degree at most r.

B.5 Numerical differentiation

In `SEYFERT` the derivatives of the angular power spectra are computed using the SteM numerical method, which has been developed by Stefano Camera [82]. This algorithm has proven to be numerically more stable than classical finite-difference methods, like the N-points stencil derivative. In this section this algorithm is briefly described, also detailing how it has been implemented in `SEYFERT`.

B.5.1 SteM algorithm

Given an independent variable x and a dependent variable y, the SteM algorithm uses an iterative linear regression of N samples of y against N samples of x to give a numerical estimate of the derivative

$$\left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{x=x_0} \tag{B.20}$$

The x variable is sampled around the central value x_0 in a symmetric way. The displaced values are computed using displacements relative[2] to the value of x_0

$$x_i = x_0 + \epsilon_i \cdot x_0 = x_0(1 + \epsilon_i) \,, \tag{B.21}$$

and the function $y(x)$ is sampled accordingly in these points

$$y_i = y(x_i) \,. \tag{B.22}$$

In `SEYFERT` a total of 15 samples is employed: the central value, 7 displacements on the right, and the 7 symmetric ones on the left. The values of the 7 positive displacements used are the following

$$\epsilon_i = \{0.00625, 0.0125, 0.01875, 0.025, 0.0375, 0.05, 0.1\} \tag{B.23}$$

accompanied by their symmetric negative ones. Therefore, the angular power spectra must be computed for 15 values of a given free parameter in order to compute the derivative with respect to it. Each parameter is varied independently of the others; if N_{params} are let free to vary this leads to a total of

$$N_{\mathrm{samples}} = 1 + 14 \cdot N_{\mathrm{params}} \tag{B.24}$$

[2] If $x_0 = 0$, as in the case of w_a, a reference value of $x_0 = 1$ is used to compute the displacements.

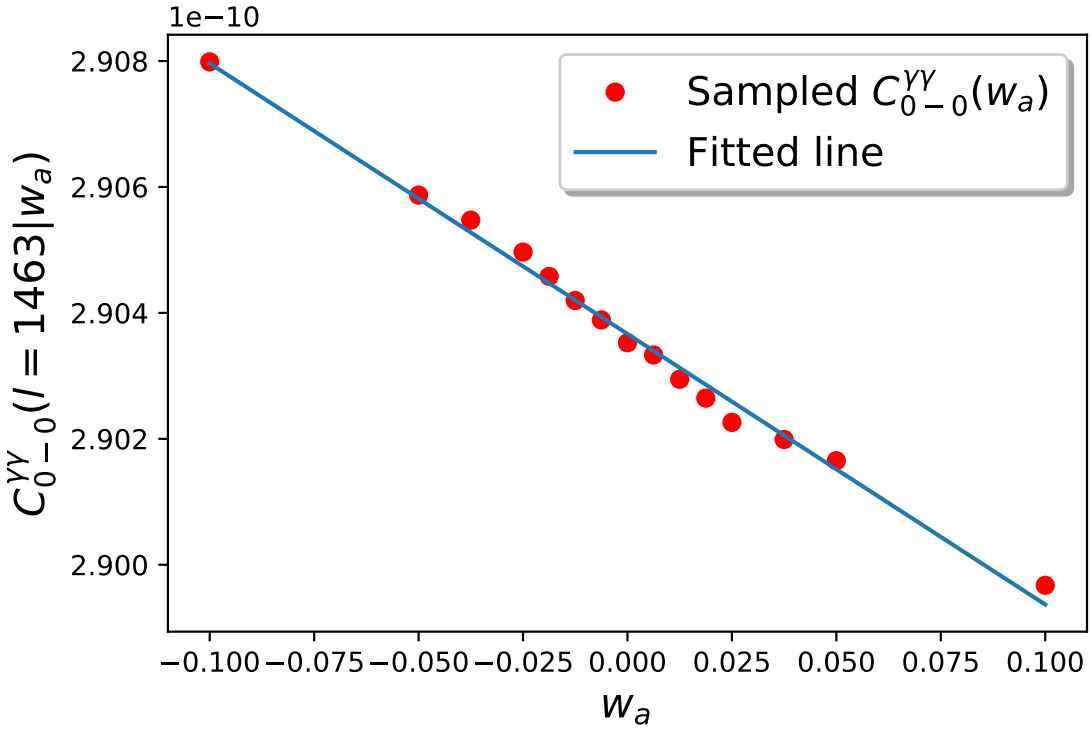

Figure B.6: Visualization of the SteM linear fit for the computation of the derivative of a particular value of the WL (denoted by γ) auto-correlation. In this example the varying parameter is w_a, whose value in the reference cosmology is 0. Therefore the displacements are computed using 1 as reference value.

evaluations of the $C(\ell)$'s. In a typical run there are 7 free cosmological parameters and 17 nuisance parameters – 10 bias parameters for $\mathrm{GC_{ph}}$, 4 for $\mathrm{GC_{sp}}$ and the 3 intrinsic alignment parameters for WL – leading to a total of 337 evaluations.

Once the x and y samples are given an iterative procedure is started: at each step a linear fit of y versus x is done, evaluating the maximum absolute percentage deviation of the y values from the fitting straight line. If this deviation is smaller than 1% the algorithm stops, returning the slope of the line as the value of the derivative. If the deviation is larger, the first and the last values of x and y are removed, going to the next iteration. If only 3 samples are left from the previous iterations, the algorithm stops independently of the value of the deviation, since a linear fit with less than 3 points does not make any sense.

Regarding the stability of the SteM algorithm, numerical checks had been done in [5] already. In particular, the SteM algorithm proved to be more stable than standard finite-difference methods. For this reason in this work no extensive tests have been repeated about the stability of the numerical differentiation.

B.5.2 Vectorisation of the linear regression

In order to compute the derivatives of the angular power spectra, a separate linear fit must be done for each combination of ℓ, i, j. As explained in Sec. B.1, the input for the computation

is a 4D `numpy` array `c_dvar_lij`: the first dimension represent the parameter with respect to which the differentiation is done, while the other three are ℓ, i, j. The first and simplest implementation one might come out with is to write a triple nested `for` loop over the last three dimensions of the array ℓ, i, j. However, since Python is an interpreted language, this approach may be quite slow. Instead, in `SEYFERT` the computation is vectorised by exploiting the efficiency of the `numpy.polyfit` function. This function accepts two arrays as inputs: `x` and `y`, respectively representing the x and y values to be fitted. Also the order of the polynomial fitting must be specified, which is 1 in this case for a linear fit. The `x` array must be one-dimensional, while `y` can be at most two-dimensional. The first dimension of `y` must match the length of `x`, while the second is independent. The function performs a polynomial fit of the columns of `y` against `x`, and it is about 100 times faster than an ordinary python `for` loop. In practice, in `SEYFERT` the 4D array `c_dvar_lij` is reshaped in 2D form using the `numpy.reshape` function, grouping the last three dimensions into one. Then the reshaped array is passed to `polyfit` along with the 15 values of the varied parameter as `x`, and after the computation an inverse reshaping is done to get the 3D `dc_lij` array of the derivative.

Bibliography

[1] Adam G. Riess et al. "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant". In: *The Astronomical Journal* (1998). ISSN: 00046256. DOI: 10.1086/300499.

[2] S. Perlmutter et al. "Measurements of Ω and Λ from 42 High-Redshift Supernovae". In: *The Astrophysical Journal* 517.2 (1999), pp. 565–586. ISSN: 0004-637X. DOI: 10.1086/307221. arXiv: 9812133v1 [astro-ph].

[3] P. A.R. Ade et al. "Planck 2013 results. I. Overview of products and scientific results". In: *Astronomy and Astrophysics* (2014). ISSN: 14320746. DOI: 10.1051/0004-6361/201321529.

[4] N. Aghanim. "Planck 2018 results. VI. Cosmological parameters". In: *Astronomy & Astrophysics* 63 (2020). ISSN: 0004-6361. DOI: 10.1051/0004-6361/201833910. URL: https://pla.esac.esa.int.

[5] A. Blanchard et al. "Euclid preparation: VII. Forecast validation for Euclid cosmological probes". In: *Astronomy & Astrophysics* 191 (2020), pp. 1–66.

[6] Scott Dodelson and Fabian Schmidt. *Modern cosmology*. II. Elsevier, 2021. ISBN: 9780128159484. DOI: 10.1016/b978-0-12-815948-4.00007-3.

[7] Sean Carroll. *Spacetime and Geometry*. Addison-Wesley, 2004.

[8] R. S. Gonçalves, G. C. Carvalho, C. A.P. P Bengaly, J. C. Carvalho, and J. S. Alcaniz. "Measuring the scale of cosmic homogeneity with SDSS-IV DR14 quasars". In: *Monthly Notices of the Royal Astronomical Society* 481.4 (Dec. 2018), pp. 5270–5274. ISSN: 13652966. DOI: 10.1093/MNRAS/STY2670. URL: https://academic.oup.com/mnras/article/481/4/5270/5114592.

[9] D. J. Fixsen. "The temperature of the cosmic microwave background". In: *Astrophysical Journal* 707.2 (2009), pp. 916–920. ISSN: 15384357. DOI: 10.1088/0004-637X/707/2/916.

[10] V. Bonvin et al. "H0LiCOW - V. New COSMOGRAIL time delays of HE 0435-1223: H0 to 3.8 per cent precision from strong lensing in a flat ΛCDM model". In: *Monthly Notices of the Royal Astronomical Society* 465.4 (2017), pp. 4914–4930. ISSN: 13652966. DOI: 10.1093/mnras/stw3006.

[11] Michael R. Blanton et al. "Sloan Digital Sky Survey IV: Mapping the milky way, nearby galaxies, and the distant universe". In: *arXiv* 28 (2017). ISSN: 23318422. DOI: 10.3847/1538-3881/aa7567.

[12] Y. Fukuda et al. "Evidence for oscillation of atmospheric neutrinos". In: *Physical Review Letters* (1998). ISSN: 10797114. DOI: 10.1103/PhysRevLett.81.1562.

[13] Ivan Esteban, M. C. Gonzalez-Garcia, Alvaro Hernandez-Cabezudo, Michele Maltoni, and Thomas Schwetz. "Global analysis of three-flavour neutrino oscillations: synergies and tensions in the determination of θ 23 , δ CP , and the mass ordering". In: *Journal of High Energy Physics* 2019.1 (2019). ISSN: 10298479. DOI: 10.1007/JHEP01(2019)106.

[14] M. Aker et al. "Improved Upper Limit on the Neutrino Mass from a Direct Kinematic Method by KATRIN". In: *Physical Review Letters* 123.22 (2019), p. 221802. ISSN: 10797114. DOI: 10.1103/PhysRevLett.123.221802. URL: https://doi.org/10.1103/PhysRevLett.123.221802.

[15] Alan H. Guth. "Inflationary universe: A possible solution to the horizon and flatness problems". In: *Physical Review D* (1981). ISSN: 05562821. DOI: 10.1103/PhysRevD.23.347.

[16] A. D. Linde. "A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems". In: *Physics Letters B* (1982). ISSN: 03702693. DOI: 10.1016/0370-2693(82)91219-9.

[17] Shinji Tsujikawa. "Introductory review of cosmic inflation". In: (2003), pp. 1–25. URL: http://arxiv.org/abs/hep-ph/0304257.

[18] Luca Amendola et al. "Cosmology and fundamental physics with the Euclid satellite". In: *Living Reviews in Relativity* 21.1 (2018). ISSN: 14338351. DOI: 10.1007/s41114-017-0010-3.

[19] Antony Lewis, Anthony Challinor, and Anthony Lasenby. "Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models". In: *The Astrophysical Journal* 538.2 (2000), pp. 473–476. ISSN: 0004-637X. DOI: 10.1086/309179.

[20] Julien Lesgourgues. "The Cosmic Linear Anisotropy Solving System (CLASS) III: Comparision with CAMB for LambdaCDM". In: (2011). URL: http://arxiv.org/abs/1104.2934.

[21] R. E. Smith, J. A. Peacock, A. Jenkins, S. D.M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M.P. Couchman. "Stable clustering, the halo model and non-linear cosmological power spectra". In: *Monthly Notices of the Royal Astronomical Society* (2003). ISSN: 00358711. DOI: 10.1046/j.1365-8711.2003.06503.x.

[22] A. A. Penzias and R. W. Wilson. "A Measurement of Excess Antenna Temperature at 4080 Mc/s." In: *The Astrophysical Journal* (1965). ISSN: 0004-637X. DOI: 10.1086/148307.

[23] R. H. Dicke, P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson. "Cosmic Black-Body Radiation." In: *The Astrophysical Journal* (1965). ISSN: 0004-637X. DOI: 10.1086/148306.

[24] A. Kogut et al. "COBE Differential Microwave Radiometers - Preliminary systematic error analysis". In: *The Astrophysical Journal* (1992). ISSN: 0004-637X. DOI: 10.1086/172033.

[25] D. N. Spergel et al. "First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters". In: *The Astrophysical Journal Supplement Series* (2003). ISSN: 0067-0049. DOI: 10.1086/377226.

[26] A. Abergel et al. "Planck 2013 results. XI. All-sky model of thermal dust emission". In: *Astronomy and Astrophysics* (2014). ISSN: 14320746. DOI: 10.1051/0004-6361/201323195.

[27] David Branch and G. A. Tammann. "Type Ia supernovae as standard candles". In: *Annual Review of Astronomy and Astrophysics* (1992). ISSN: 00664146. DOI: 10.1146/annurev.aa.30.090192.002043.

[28] Adam G. Riess, Stefano Casertano, Wenlong Yuan, Lucas M. Macri, and Dan Scolnic. *Large magellanic cloud cepheid standards provide a 1% foundation for the determination of the hubble constant and stronger evidence for physics beyond ΛCDM.* 2019. DOI: 10.3847/1538-4357/ab1422.

[29] Valérie De Lapparent. "A slice of the Universe". In: (1985).

[30] John Tonry and Marc Davis. "A Survey of galaxy redshifts I. Data Reduction Techniques". In: *The Astronomical Journal* 84 (1979), pp. 1511–1525.

[31] Kyle S. Dawson et al. "The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: overview and early data". In: *The Astronomical Journal* (2016). ISSN: 0004-6256. DOI: 10.3847/0004-6256/151/2/44.

[32] Daniel J. Eisenstein et al. "Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies". In: *The Astrophysical Journal* 633.2 (2005), pp. 560–574. ISSN: 0004-637X. DOI: 10.1086/466512.

[33] Martin Kilbinger. "Cosmology with cosmic shear observations: a review". In: *Reports on Progress in Physics* 78.8 (Nov. 2014). DOI: 10.1088/0034-4885/78/8/086901. URL: http://arxiv.org/abs/1411.0115%20http://dx.doi.org/10.1088/0034-4885/78/8/086901.

[34] Richard Massey et al. "COSMOS: Three-dimensional Weak Lensing and the Growth of Structure". In: *The Astrophysical Journal Supplement Series* 172.1 (2007), pp. 239–253. ISSN: 0067-0049. DOI: 10.1086/516599.

[35] David J. Bacon, Alexandre R. Refregier, and Richard S. Ellis. "Detection of weak gravitational lensing by large-scale structure". In: *Monthly Notices of the Royal Astronomical Society* 318.2 (2000), pp. 625–640. ISSN: 00358711. DOI: 10.1046/j.1365-8711.2000.03851.x.

[36] David M. Wittman, J. Anthony Tyson, David Kirkman, Ian Dell'Antonio, and Gary Bernstein. "Detection of weak gravitational lensing distortions of distant galaxies by cosmic dark matter at large scales". In: *Nature* 405.6783 (2000), pp. 143–148. ISSN: 00280836. DOI: 10.1038/35012001.

[37] L. Van Waerbeke et al. "Detection of correlated galaxy ellipticities from CFHT data: First evidence for gravitational lensing by large-scale structures". In: *Astronomy and Astrophysics* 358.1 (2000), pp. 30–44. ISSN: 00046361.

[38] Catherine Heymans et al. "CFHTLenS: The Canada-France-Hawaii Telescope Lensing Survey". In: *Monthly Notices of the Royal Astronomical Society* 427.1 (2012), pp. 146–166. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2012.21952.x.

[39] Catherine Heymans et al. "CFHTLenS tomographic weak lensing cosmological parameter constraints: Mitigating the impact of intrinsic galaxy alignments". In: *Monthly Notices of the Royal Astronomical Society* 432.3 (2013), pp. 2433–2453. ISSN: 00358711. DOI: 10.1093/mnras/stt601.

[40] Niall MacCrann, Joe Zuntz, Sarah Bridle, Bhuvnesh Jain, and Matthew R. Becker. "Cosmic discordance: Are Planck CMB and CFHTLenS weak lensing measurements out of tune?" In: *Monthly Notices of the Royal Astronomical Society* 451.3 (2015), pp. 2877–2888. ISSN: 13652966. DOI: 10.1093/mnras/stv1154.

[41] T. M.C. Abbott et al. "Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing". In: *Physical Review D* 98.4 (2018), pp. 1–31. ISSN: 24700029. DOI: 10.1103/PhysRevD.98.043526.

[42] David H. Weinberg, Michael J. Mortonson, Daniel J. Eisenstein, Christopher Hirata, Adam G. Riess, and Eduardo Rozo. "Observational probes of cosmic acceleration". In: *Physics Reports* 530.2 (Sept. 2013), pp. 87–255. ISSN: 03701573. DOI: 10.1016/j.physrep.2013.05.001.

[43] Michel Chevallier and David Polarski. "Accelerating universes with scaling dark matter". In: *International Journal of Modern Physics D* 10.2 (2001), pp. 213–223. ISSN: 02182718. DOI: 10.1142/S0218271801000822.

[44] Eric V. Linder. "Exploring the Expansion History of the Universe". In: *Physical Review Letters* 90.9 (2003), p. 4. ISSN: 10797114. DOI: 10.1103/PhysRevLett.90.091301.

[45] Giuseppe D. Racca et al. "The Euclid mission design". In: *Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave* 9904.July 2016 (2016), 99040O. ISSN: 1996756X. DOI: 10.1117/12.2230762.

[46] R Laureijs, J Amiaux, S Arduini, J-L Augueres, J Brinchmann, R Cole, M Cropper, C Dabin, L Duvet, and A Ealet. "Euclid Redbook". In: *arXiv preprint arXiv:1110.3193* (2011).

[47] H. Meusinger et al. "Exploring the Expansion History of the Universe". In: *Astronomy and Astrophysics* 522.2 (2012), pp. 1–16. ISSN: 10797114. DOI: 10.1103/PhysRevLett.90.091301.

[48] Federico Fornari. "Software Development and Detector Characterization of the EUCLID Near-Infrared Spectro-Photometer". PhD thesis. 2018.

[49] Mark S. Cropper et al. "VIS: the visible imager for Euclid". In: 2018. ISBN: 9781510619494. DOI: 10.1117/12.2315372.

[50] Thierry Maciaszek et al. "Euclid Near Infrared Spectrometer and Photometer instrument concept and first test results obtained for different breadboards models at the end of phase C". In: *Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave* 9904 (2016), 99040T. ISSN: 1996756X. DOI: 10.1117/12.2232941.

[51] M. Scodeggio et al. "The VVDS Data-Reduction Pipeline: Introducing VIPGI, the VIMOS Interactive Pipeline and Graphical Interface". In: *Publications of the Astronomical Society of the Pacific* 117.837 (2005), pp. 1284–1295. ISSN: 0004-6280. DOI: 10.1086/496937. arXiv: 0409248v1 [astro-ph].

[52] Paul Jerram and James Beletic. "Teledyne's high performance infrared detectors for Space missions". In: 11180.October 2018 (2019), p. 120. ISSN: 1996756X. DOI: 10.1117/12.2536040.

[53] Yannick Copin, Institut De Physique, and Infinis Université. *SIR : Slitless Spectroscopy in Euclid Slitless spectroscopy in Euclid*. 2020.

[54] Marco Scodeggio. *Euclid SGS SIR Requirements Specification Document*. 2015.

[55] Marco Scodeggio. *Euclid SGS SIR Processing Function Validation Plan*. 2015.

[56] Santiago Serrano. *Euclid Simulations*. Les Houches, 2020.

[57] T. Prusti et al. "The Gaia mission". In: *Astronomy and Astrophysics* 595.29272 (2016). ISSN: 14320746. DOI: 10.1051/0004-6361/201629272.

[58] A. Robin and M. Crézé. "Stellar populations in the Milky Way: a synthetic model". In: *Astronomy & Astrophysics* (1986).

[59] A. J. Pickles. "A Stellar Spectral Flux Library: 1150–25000 Å". In: *Publications of the Astronomical Society of the Pacific* 110.749 (1998), pp. 863–878. ISSN: 0004-6280. DOI: 10.1086/316197.

[60] R. Drimmel, A. Cabrera-Lavers, and M. López-Corredoira. "A three-dimensional Galactic extinction model". In: *Astronomy & Astrophysics* 409.1 (Oct. 2003), pp. 205–215. ISSN: 0004-6361. DOI: 10.1051/0004-6361:20031070. URL: http://www.aanda.org/10.1051/0004-6361:20031070.

[61] Th Lejeune, F. Cuisinier, and R. Buser. "A standard stellar library for evolutionary synthesis: I. Calibration of theoretical spectra". In: *Astronomy and Astrophysics Supplement Series* 125.2 (Oct. 1997), pp. 229–246. ISSN: 03650138. DOI: 10.1051/aas:1997373. URL: https://ui.adsabs.harvard.edu/abs/1997A&AS..125..229L/abstract.

[62] Douglas Potter, Joachim Stadel, and Romain Teyssier. "PKDGRAV3: beyond trillion particle cosmological simulations for the next era of galaxy surveys". In: *Computational Astrophysics and Cosmology* 4.1 (2017). ISSN: 2197-7909. DOI: 10.1186/s40668-017-0021-1.

[63] L. Miller et al. "Bayesian galaxy shape measurement for weak lensing surveys - III. Application to the Canada-France-Hawaii Telescope Lensing Survey". In: *Monthly Notices of the Royal Astronomical Society* 429.4 (Mar. 2013), pp. 2858–2880. ISSN: 00358711. DOI: 10.1093/mnras/sts454. URL: www.cfhtlens.org.

[64] B. T.P. Rowe et al. "GalSim: The modular galaxy image simulation toolkit". In: *Astronomy and Computing* 10 (Apr. 2015), pp. 121–150. ISSN: 22131337. DOI: 10.1016/j.ascom.2015.02.002. URL: http://wfirst.gsfc.nasa.gov/.

[65] N. Scoville et al. "The Cosmic Evolution Survey (COSMOS): Overview". In: *The Astrophysical Journal Supplement Series* 172.1 (2007), pp. 1–8. ISSN: 0067-0049. DOI: 10.1086/516585.

[66] R. B. Tully and J. R. Fisher. *A new method of determining distance to galaxies*. Tech. rep. 1977, pp. 105–117. URL: https://ui.adsabs.harvard.edu/abs/1977A&A....54..661T/abstract.

[67] 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 Hα emitters for future spectroscopic near-IR space missions". In: *Astronomy and Astrophysics* 590 (2016), pp. 1–16. ISSN: 14320746. DOI: 10.1051/0004-6361/201527081.

[68] M. Kümmel, J. R. Walsh, N. Pirzkal, H. Kuntschner, and A. Pasquali. "The Slitless Spectroscopy Data Extraction Software aXe". In: *Publications of the Astronomical Society of the Pacific* 121.875 (2009), pp. 59–72. ISSN: 0004-6280. DOI: 10.1086/596715.

[69] M. Scodeggio and C. Mancini. *Results of the validation of the SIR software pipeline for the extraction of Euclid near-infrared spectra*. Private communication. 2020.

[70] Louis Gabarra. *A key tool to probe Euclid spectroscopy: Spectro-Photometric simulations of galaxies to unravel NISP's capabilities*. Sept. 2020. URL: http://tesi.cab.unipd.it/64917/.

[71] *IST:F Public Results: Fisher matrices*. URL: https://github.com/euclidist-forecasting/fisher_for_public.

[72] Ryuichi Takahashi, Masanori Sato, Takahiro Nishimichi, Atsushi Taruya, and Masamune Oguri. "Revising the halofit model for the nonlinear matter power spectrum". In: *The Astrophysical Journal* 761.10pp (2012), p. 152. DOI: 10.1088/0004-637X/761/2/152.

[73] *IST:F Public Results: Power Spectra*. URL: https://u.pcloud.link/publink/show?code=0wJotalK#folder=2895054161&tpl=publicfoldergrid.

[74] Benjamin R. Granett. *Results of the Euclid Science Performance Verification 2*. Private communication. 2020.

[75] Martin Kilbinger. "Cosmology with cosmic shear observations: A review". In: *Reports on Progress in Physics* 78.8 (2015). ISSN: 00344885. DOI: 10.1088/0034-4885/78/8/086901.

[76] Thomas D. Kitching, Justin Alsing, Alan F. Heavens, Raul Jimenez, Jason D. McEwen, and Licia Verde. "The limits of cosmic shear". In: *Monthly Notices of the Royal Astronomical Society* 469.3 (Aug. 2017), pp. 2737–2749. ISSN: 13652966. DOI: 10.1093/mnras/stx1039. URL: http://euclid-ec.org.

[77] Pablo Lemos, Anthony Challinor, and George Efstathiou. "The effect of Limber and flat-sky approximations on galaxy weak lensing". In: *Journal of Cosmology and Astroparticle Physics* 2017.5 (May 2017), p. 014. ISSN: 14757516. DOI: 10.1088/1475-7516/2017/05/014. URL: https://iopscience.iop.org/article/10.1088/1475-7516/2017/05/014%20https://iopscience.iop.org/article/10.1088/1475-7516/2017/05/014/meta.

[78] Benjamin Joachimi et al. *Galaxy Alignments: An Overview*. Nov. 2015. DOI: 10.1007/s11214-015-0177-4. URL: https://link.springer.com/article/10.1007/s11214-015-0177-4.

[79] Alina Kiessling et al. "Galaxy Alignments: Theory, Modelling & Simulations". In: *Space Science Reviews* 193.1-4 (2015), pp. 67–136. ISSN: 15729672. DOI: 10.1007/s11214-015-0203-6.

[80] Donnacha Kirk et al. "Galaxy alignments: Observations and impact on cosmology". In: *Space Science Reviews* 193.1-4 (Apr. 2015), pp. 139–211. DOI: 10.1007/s11214-015-0213-4. URL: http://arxiv.org/abs/1504.05465%20http://dx.doi.org/10.1007/s11214-015-0213-4.

[81] Sarah Bridle and Lindsay King. "On photometric redshift requirements". In: *New Journal of Physics* 9 (2007). ISSN: 13672630. DOI: 10.1088/1367-2630/9/12/444.

[82] Stefano Camera, Ian Harrison, Anna Bonaldi, and Michael L Brown. "SKA weak lensing-III. Added value of multiwavelength synergies for the mitigation of systematics". In: *MNRAS* 464 (2017), pp. 4747–4760. DOI: 10.1093/mnras/stw2688. URL: http://skatelescope.org.

[83] Nagar D. K Gupta A. K. *Matrix Variate Distributions*. Chapman & Hall, 2000.

[84] Tobias Baldauf, Uroš Seljak, Robert E. Smith, Nico Hamaus, and Vincent Desjacques. "Halo stochasticity from exclusion and nonlinear clustering". In: *Physical Review D - Particles, Fields, Gravitation and Cosmology* 88.8 (2013), pp. 1–27. ISSN: 15507998. DOI: 10.1103/PhysRevD.88.083507.

[85] Catherine Heymans et al. "KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints". In: *arXiv* (2020). ISSN: 0004-6361. DOI: 10.1051/0004-6361/202039063.

[86] Xiao Fang, Elisabeth Krause, Tim Eifler, and Niall MacCrann. "Beyond Limber: Efficient computation of angular power spectra for galaxy clustering and weak lensing". In: *Journal of Cosmology and Astroparticle Physics* 2020.5 (2020). ISSN: 14757516. DOI: 10.1088/1475-7516/2020/05/010.

[87] E. Krause et al. "Dark Energy Survey Year 1 Results: Multi-Probe Methodology and Simulated Likelihood Analyses". In: (2017), pp. 1–22. ISSN: 2331-8422. URL: http://arxiv.org/abs/1706.09359.

[88] Fabien Lacasa and Julien Grain. "Fast and easy super-sample covariance of large-scale structure observables". In: *Astronomy and Astrophysics* 624 (2019), pp. 1–14. ISSN: 14320746. DOI: 10.1051/0004-6361/201834343.

[89] Samira Hamimeche and Antony Lewis. "Likelihood analysis of CMB temperature and polarization power spectra". In: *Physical Review D - Particles, Fields, Gravitation and Cosmology* 77.10 (2008), pp. 1–35. ISSN: 15507998. DOI: 10.1103/PhysRevD.77.103013.

[90] *IBM Spectrum LSF Suites*. Web page. URL: https://www.ibm.com/products/hpc-workload-management.

[91] Åke Bjorck Germund Dahlquist. *Numerical Methods in Scientific Computing*. Vol. 1. Philadelphia: Society for Industrial and Applied Mathematics, 2008. ISBN: 9780898716443.

[92] A Quarteroni, R Sacco, and F Saleri. *Numerical Mathematics*. 2007, p. 655. ISBN: 9783540346586.