Alma Mater Studiorum · Università di Bologna

Scuola di Scienze
Dipartimento di Fisica e Astronomia "Augusto Righi"
Corso di Laurea magistrale in Astrofisica e Cosmologia

AMICO-WL: an algorithm for weak-lensing detection of galaxy clusters

Relatore:
Prof. Lauro Moscardini

Presentata da:
Leonardo Trobbiani

Corelatori:
Priv-Doz. Dr. Matteo Maturi
Dr. Carlo Giocoli

Anno Accademico 2021/2022

Abstract

We developed a highly optimized routine for the detection of clusters of galaxies: AMICO-WL, that implements an improved version of the optimal linear matched filter presented in Maturi et al. (2005) inside the computational environment of AMICO (Bellagamba et al., 2018). The filter is optimized to maximize the signal-to-noise ratio of the detections and minimize the number of spurious detections caused by the superposition of large-scale structures.

Our goal is to detect the matter distribution of clusters, represented mainly by their dark matter halos, through their weak gravitational lensing signatures by applying the filter to a catalog of simulated galaxy ellipticities. The simulated data represents the dataset quality expected for the *Euclid* mission, we thus also provide a forecast for the filter performances on weak lensing for this mission.

We investigated the sample in terms of *completeness* and *purity* by matching the detections with the halos in the simulations. We found $SNR \sim 2.5$ as a reliable threshold for the detection, in fact the purity is close to the "expected purity", 70% against the expected 80%. We studied the completeness in the redshift-mass plane $z - M_{200}$ of the halos. The completeness reaches almost 100% in the case of the halos with virial mass $\log_{10}(M_{200}/M_{\odot}/h) > 14.4$ and redshift in the range [0.2,0.4]. As expected we reach the higher completeness where the efficiency of the gravitational lensing effect is maximum, i.e where the lens is located in between the sources and the observer.

We further refined the strategy by using only the galaxies with $z > 0.6$ to remove most of the non-lensed foreground sources then compared the results with the analysis performed with the full galaxy catalog. Using the truncated dataset we detected 607 more objects and a consequent higher completeness for the truncated catalog with respect to the complete one, 47% against 45% for a fixed 65% purity in booth catalogs. Using the truncated catalog we also increment the sensitivity to halos with smaller masses.

The ultimate goal is to build an optimized routine for the weak-lensing detection of clusters that will work in complementary and possibly in combination with the photometric detection algorithm implemented in AMICO for the analysis of KiDS, CHFTLens, DES, and the upcoming LSST and Euclid.

Keywords: cosmology:theory,dark matter, gravitational lensing

Abstract

In questa tesi abbiamo sviluppato duna routine altamente ottimizzata per il rilevamento di ammassi di galassie: AMICO-WL, che implementa una nuova versione del *filtro adattivo lineare ottimale* presentato in Maturi et al. (2005). Il codice è stato sviluppato all'interno dell'ambiente computazionale dell'algoritmo AMICO (Bellagamba et al., 2018).

Il nostro obiettivo è rilevare la distribuzione di materia degli ammassi di galassie, rappresentata principalmente dai loro aloni di materia oscura, attraverso gli effetti di weak-lensing applicando il filtro a un catalogo di ellitticità galattiche simulate. I dati simulati rappresentano le caratteristiche dei dataset attese per la missione *Euclid*, forniamo quindi anche una previsione delle prestazioni del codice per questa missione.

Abbiamo investigato il campione in termini di *completezza* e *purezza* confrontando le detezioni con gli ammassi simulati. Abbiamo trovato che un valore di $SNR \sim 2.5$ costituisce una soglia affidabile per la rilevazione, infatti la purezza è vicina alla "purezza attesa", pari al 70% rispetto all'80% atteso. Abbiamo studiato la completezza nel piano redshift-massa $z - M_{200}$ degli ammassi. La completezza raggiunge quasi il 100% nel caso degli ammassi con massa viriale $\log_{10}(M_{200}/M_{\odot}/h) > 14.4$ e redshift nell'intervallo [0.2,0.4]. Come ci si aspetta, si raggiunge la maggiore completezza dove l'efficienza dell'effetto di lente gravitazionale è massima, ovvero dove la lente si trova tra le sorgenti e l'osservatore.

Abbiamo ulteriormente affinato la strategia utilizzando solo le galassie con $z > 0,6$ per rimuovere la maggior parte delle sorgenti di foreground non lensate, quindi abbiamo confrontato i risultati con l'analisi eseguita con l'intero catalogo di galassie. Utilizzando il dataset troncato abbiamo rilevato 607 oggetti in più e ottenuto una conseguente maggiore completezza per il catalogo troncato rispetto a quello completo, del 47% rispetto al 45% per una purezza fissa del 65% in entrambi i cataloghi. Utilizzando il catalogo troncato abbiamo anche aumentato la sensibilità agli ammassi con masse più piccole.

L'obiettivo è costruire una routine ottimizzata per la detezione di ammassi tramite weak-lensing che agirà in modo complementare e possibilmente in combinazione con per la detezione in fotometria implementato in AMICO per l'analisi di KiDS, CHFTLens, DES e i prossimi progetti LSST ed Euclid.

Keywords: cosmologia:teoria, materia oscura, lensing gravitazionale

Contents

Introduction **1**

1 The ΛCDM model and structure formation theory **3**
1.1 Cosmological framework . . . 3
1.1.1 The ΛCDM model . . . 7
1.2 Density perturbations and their evolution . . . 9
1.2.1 Statistical tools for Cosmology . . . 12
1.2.2 The cosmological power spectrum . . . 13

2 Gravitational lensing **17**
2.1 Lensing theory . . . 18
2.1.1 First-order lens mapping and lensing regimes . . . 20
2.1.2 Properties of the shear tensor . . . 22
2.1.3 E and B mode decomposition . . . 24
2.2 A weak lensing estimator: the aperture mass . . . 25
2.2.1 Tangential and cross shear components . . . 28
2.2.2 Observable galaxy ellipticities . . . 29

3 Galaxy clusters: detection and their use in comsology **31**
3.1 Cluster cosmology . . . 32
3.2 Cluster detection . . . 34
3.2.1 Optical and NIR observations . . . 34
3.2.2 X-ray observations . . . 34
3.2.3 Sunyaev Zel'dovich effect . . . 35
3.2.4 Gravitational lensing . . . 35

4 The implementation of an optimal filter to weak lensing: AMICO-WL **37**
4.1 The theory of the optimal filter . . . 38
4.1.1 Definition of the optimal filter . . . 39
4.1.2 Noise properties . . . 39
4.1.3 Halo lensing signal model . . . 41
4.2 Augmenting AMICO for weak-gravitational lensing measures . . . 42
4.2.1 Input of AMICO-WL . . . 44
4.2.2 Creation of the weak-lensing signal maps . . . 45
4.2.3 Detection algorithm and cleaning procedure for the weak-lensing case . . . 46

5 Analysis of the Euclid simulations with AMICO-WL **51**
5.1 The Euclid mission 51
5.2 Simulated Euclid catalogs 53
5.2.1 The numerical simulations 53
5.2.2 The simulated data 54
5.3 Application of AMICO-WL to the mock catalogs 56
5.3.1 Criteria for the choice of the detection threshold 57
5.3.2 Catalog of detections 64
5.4 Matching procedure and weak lensing selection function 64
5.4.1 Matching analysis 70
5.4.2 Signal-to-noise ratio 71
5.4.3 Redshift and virial mass 71
5.4.4 The un-matched halos 75
5.5 Evaluation of the sample purity and completeness 79
5.5.1 Random match test 81
5.6 Effect of varying the input catalog of galaxy ellipticities 84

6 Conclusions and future perspectives **91**

A Analysis of the complete ellipticity catalog **93**

Introduction

Galaxy clusters are peculiar astrophysical objects whose detection is a crucial task in Astrophysics and Cosmology. As the most massive structures that had the time to form in the present-day Universe, clusters, and especially their *dark matter halos*, are tracers of the density peaks in the large-scale matter distribution, and therefore they are strongly coupled with the expansion history of the Universe, as well as with the formation and evolution of structures, and it is with measuring their number and spatial distribution that we can put constraints on cosmological models.
Galaxy cluster detection has been performed through the years with several successful methods: the observation of the hot and highly ionized gas clusters are embedded in, which makes them bright X-ray sources (e.g. Rosati et al. (2002)), the detection of distortions in the cosmic microwave background spectrum (Sunyaev and Zeldovich (1970); e.g. Bleem et al. (2015)) or the detection of clusters in optical and near-infrared surveys (e.g AMICO algorithm; Bellagamba et al. (2011); Bellagamba et al. (2018)).

In addition to those methods it is possible to detect galaxy clusters through the study of the gravitational lensing effect of their matter distribution, i.e. dark matter halos, on the background sources. Specifically, the detection is possible in the context of the *weak-lensing* regime, in which the deflection of light is described at the first-order approximation. Detecting dark matter halos with lensing is also important because this method is dependent only on the matter distribution most of which is dark matter. Thus it is possible to detect objects with a very low electromagnetic emission that would be hidden from the other methods.

One of the main sources of error in these measurements is the fact that dark matter distribution in the Universe is not a discrete distribution that contains only the halos, but it can be thought as a continuous distribution where the halos are peaks of mass density embedded into a lower density *large-scale structure*, the cosmic web that connects them. Measuring the inhomogeneities of projected mass distribution through lensing observations the signal of halos is overlaid by the lensing signal of the large-scale structure in front of and behind the halos. Since it is approximately a Gaussian random field, lensing by large-scale structure adds peaks and troughs to the signal which can be mistaken for halos, so part of the detections of candidates of dark matter halos can actually be maxima in the random weak gravitational lensing signal of the large-scale structure.

These two types of signal are not possible to be strictly separated because there is not a sharp boundary between the halos and the large-scale structure, but it is possible to define a typical scale that helps to discriminate between their lensing signals. The large-scale structure can be considered composed of dark-matter halos of a broad and continuous mass range. At each cosmological epoch, it is possible to

define the *nonlinear scale*, where the variance of the dark matter density contrast becomes unity, which sets an operational definition of the lensing signal due to the halos as the signal contributed by non-linear, gravitationally bound, virialized structures; the lensing signal due to the large scale structure is instead the signal contributed by the linearly evolved matter distribution beyond the non-linear scale. It was first shown in Maturi et al. (2005) and Maturi et al. (2007) that is possible to use these concepts to construct a linear matched filter capable to detect the weak lensing signal of dark matter halos of galaxy clusters while also suppressing the signal of the large scale structure. The result is that the spurious detections attributed to random peaks in the projected mass density of large-scale structures are reduced. The filter is then searching for those halos that do create the non-linear power spectrum, while the linearly evolved structures are treated as a noise component influencing the shape of the matched filter.

The goal of this Thesis is to produce a highly optimized routine for the detection of clusters using the computational background of AMICO (Bellagamba et al. (2011); Bellagamba et al. (2018)), an algorithm that implements an optimal matched filter for optical detection of clusters and adapting it for weak lensing detection of dark matter halos in wide field surveys data. We want also to test the code with a simulated ellipticity catalog, the observable of the weak-lensing, with the characteristics of a wide field survey performed by the future ESA *Euclid* mission. The outline of this Thesis is:

- Chapter 1 will give an introduction to the cosmological background and elements of the theory of structure formation. The goal is to give a brief overview to provide the reader with the tools to follow the discussion of the following chapters;
- Chapter 2 will give the fundamentals of the theory of gravitational lensing and the theoretical description of the *aperture mass*, the estimator of weak-lensing signal used to detect halos;
- Chapter 3 will contain a description of the use of clusters of galaxies in Cosmology and different methods to detect them in various observational datasets;
- Chapter 4 will contain the description of **AMICO-WL**, the implemented code, based on the theory of the optimal filter;
- Chapter 5 we will give the results of the AMICO-WL tests performed on Euclid simulated data in terms of *completeness* and *purity*, used to define the quality of an extracted sample of cluster candidates;

We will end the Thesis with the final Chapter 6 where we review and summarise the main results of this Thesis, and discuss their relevance and the possible impact on future works.

Chapter 1

The ΛCDM model and structure formation theory

This Chapter contains an overview of the standard model of Cosmology and the essential concepts that will be useful in the following chapters. The standard model consists of a modelization for the cosmological background, which is a homogeneous and isotropic solution of the field equations of General Relativity, and a theory for structure formation. The background model is depicted by the Friedmann-Lemaître-Robertson-Walker metric, FLRW, where hypersurfaces of constant time are homogeneous and isotropic three-dimensional spaces, either flat or curved, and change with time according to a scale factor that depends on time only. Friedmann provided the two equations that determine the dynamics of the scale factor, they follow from Einstein's field equations given the highly symmetric form of the metric. Crucial to this Thesis is the evolution and the formation of structures, in particular clusters, which current theories assume to be structures grown through gravitational instability from initial seed perturbations, probably originating from cosmological inflation. The inflation theory suggests that the statistics of the seed fluctuations in the primordial density field are almost Gaussian. It is possible to use linear perturbation theory for most of the evolutionary history of the fluctuations as their amplitude is low until the late stages.

We will start in Section 1.1 with an introduction of the essential concepts needed to construct a cosmological framework, and then we will focus on the currently favored cosmological model, the so-called ΛCDM model, in Section 1.1.1 and the core part will be the description of *density perturbations*, the structures that will ultimately grow into the cosmic structures we observe today. The main references for this Chapter are Bartelmann and Schneider (2001), Schneider (2006) and Carroll (2004).

1.1 Cosmological framework

Cosmology is the branch of physics that has as its object of study the Universe as its whole, the goal is to understand the origin and evolution through theoretical and observational tools. The structure, content, and dynamic of the Universe are

described using cosmological models that are based on a set of quantities called cosmological parameters, which can be observationally constrained. In modern cosmology, the assumption on which the models are based is the so-called *cosmological principle*, which states that on scales typically larger than hundreds of Mpc the Universe is homogeneous and isotropic. What follows the homogeneity is what historically is called the *copernican principle*. There is no privileged position or direction, matter distribution and properties of the Universe look the same everywhere and in every direction. Fundamental evidence of these two principles is the observation of the cosmic microwave background (CMB), the relic radiation of the early stages of the Universe, measured for the first time by Penzias and Wilson (Penzias and Wilson, 1965) in 1965, marking a cornerstone in Cosmology. The isotropy of this radiation suggests that the principles are valid on scales larger than Mpc.

The other pillar on which modern cosmological models rely is the theory of *General Relativity* by Einstein (1916) that describes the gravitational force that dominates on large scales. It is possible to prove that from the cosmological principles and General Relativity the mathematical background of models, the metric of the Universe, is constrained in the *Friedmann-Lemaître-Robertson-Walker* (Friedmann (1922);Friedmann (1924); Robertson (1935)), FLRW form:

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - a^2(t)\left[\frac{1}{1-KR^2}\mathrm{d}R^2 + R^2(\mathrm{d}\theta^2 + \sin\theta^2\mathrm{d}\phi^2)\right], \tag{1.1}$$

that expresses an element of space-time in polar coordinates (R, θ, ϕ). We can assume them to be adimensional, c is the speed of light, t is the proper cosmic time, and $a(t)$ is the *scale factor*: in fact, the spatial metric can only isotropically contract or expand with the scale factor which must be a function of time only, because otherwise, the expansion would be different at different places, violating homogeneity. The constant $K \in \mathbb{E}$ is the *curvature constant*: if $K = 0$ the Universe has a flat geometry; if $K > 0$, the Universe is closed with a spherical geometry; if $K < 0$, the Universe is open with a hyperbolic geometry. Observational pieces of evidence support the Euclidean geometry scenario (flat Universe) (see e.g. Planck Collaboration Aghanim et al. (2020)).

Before the discovery of CMB, the main observational cornerstone of cosmology has been the Hubble-Lemaître law first described theoretically by Lemaître (1927) and then observed by Hubble (1929) on distances and velocities of a galaxy sample. This was the first sign of the expansion of the Universe: in fact, the two quantities were found to be proportional with a positive proportionality coefficient, meaning that increasing separation between galaxies implies increasing receding velocity. The coefficient of proportionality is the so-called Hubble parameter $H(t)$, whose value at the present time is referred to as H_0. We should not intend this recession of galaxies as a movement of the intrinsic position of galaxies in space but a movement of fixed objects in space due to the expansion of space itself; often, it is called *Hubble flow*. It becomes essential to use the *co-moving coordinates* and the associated co-moving distance x as they do not depend on time, unlike the physical distance $r(t)$ does. Two are linked by the scale factor as $x = r(t)/a(t)$. Therefore we have the Hubble-Lemaître law:

$$v(t) = \frac{\mathrm{d}r(t)}{\mathrm{d}t} = x\frac{\mathrm{d}a(t)}{\mathrm{d}t} = \frac{\mathrm{d}a(t)/\mathrm{d}t}{a}r(t), \tag{1.2}$$

the Hubble parameter is then defined as the time derivative of the logarithm of the scale factor:

$$H(t) \equiv \frac{\mathrm{d} \ln a(t)}{\mathrm{d}t} = \frac{\dot{a}}{a} \tag{1.3}$$

The Hubble constant, the value of the Hubble parameter at the present time, is expressed as a velocity over a distance, generally indicated with $H_0 \equiv H(t_0) = 100h$ km s^{-1} Mpc^{-1}, being h a dimensionless constant also used to express observables in a parameter-independent way, since the precise value of H_0 is still on the debate (see e.g. Freedman and Madore (2010), Schneider (2006)). The time scale for the expansion of the Universe is the *Hubble time* $H_0^{-1} \approx 10^{10}\mathrm{yr}h^{-1}$, this is usually used as a reference of the order-of-magnitude of the age of the Universe. However, all measurements confirm that it is a positive quantity, so the Universe is *expanding*. Due to the expansion of space, photons are redshifted while propagating from the source to the observer; this effect is the *cosmological redshift*. The redshift of a photon emitted with wavelength λ_{em} and observed with a wavelength λ_{obs} is given by:

$$z \equiv \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}} = \frac{1}{a(t_{em})} - 1. \tag{1.4}$$

where we use $a_0 = 1$. Redshift and scale factor mark distances and, therefore, specific cosmic times. High-redshift objects are those whose photons crossed the longest distances and were emitted the longest time ago from today. The redshift is then also an observational proxy for the distance in space and time of the sources. What we said is true only in the case of sufficiently distant extragalactic objects, since on small scales the expansion of the Universe is not likely to dominate over intrinsic proper motion.

It is possible to define two distance expressions by using the luminosity of a source or its size. Starting from the relation between measured flux F and intrinsic luminosity L of a cosmological source, we define the *luminosity distance* d_L: $\frac{L}{4\pi d_L^2}$. This distance is linked to the redshift as:

$$d_L \equiv (1+z)x, \tag{1.5}$$

where x is the comoving distance. If we consider instead the intrinsic and measured angular size of a source, l and θ, respectively, the *angular diameter distance*, d_A, can be defined as:

$$d_A = \frac{l}{\theta} : \tag{1.6}$$

this distance would take a great role in Section 2 when talking about gravitational lensing theory.
The two cosmological distances are related by the redshift as:

$$d_A = \frac{d_L}{(1+z)^2}, \tag{1.7}$$

in a way that

$$d_A = \frac{x}{1+z}. \tag{1.8}$$

To complete the general theoretical framework of modern cosmological models, we need a set of evolutionary equations for the scale factor that quantify how the space scales with time given its geometry. Describing the content components of the

Universe as a uniform "cosmic fluids" and General Relativity equations, Friedmann (1922) (1924) derived two equations describing $a(t)$[1]:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{Kc^2}{a^2}, \tag{1.9}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right), \tag{1.10}$$

where G is the gravitational constant and ρ and p are the density and the pressure of the fluid, respectively. The Friedmann equations, when combined, yield the *adiabatic equation*:

$$\dot{\rho} + \frac{3\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) = 0. \tag{1.11}$$

We can interpret this Equation as the first law of thermodynamics in the cosmological context. Equations 1.9, 1.10 and 1.11 are not independent; thus, two of them are sufficient for the description of the dynamics of $a(t)$.

When Friedmann deduced these equations, the standard model involved a static Universe; however, the static hypothesis, together with Eq. 1.9, leads to the *unphysical* implication that density and pressure must have different signs. Rejecting the possibility of a Universe in expansion, Einstein introduced in 1916 a term Λ, to modify his gravity theory. The expansion was then confirmed with the observations of the galaxy recession, and the Λ constant was discarded. Only in recent times, it was re-introduced as representative of a new component of the Universe that would explain its accelerated expansion, as first proved by distant supernova observations (Perlmutter et al. (1999); Riess et al. (1998)). In modern models, the cosmological constant Λ is associated with a form of *dark energy*, whose nature is still unknown.

The first Friedmann Eq. 1.9 can be written in a different form if we use the definition of $H(t)$ and as suggested we assume flat geometry ($K = 0$):

$$H^2(t) = \frac{8\pi G}{3}\rho, \tag{1.12}$$

which by isolating the mass density, gives the expression for the so-called *critical density*

$$\rho_{crit} \equiv \frac{3H^2(t)}{8\pi G}, \tag{1.13}$$

which has a present-time value of $\sim 10^{-29} h^2 \mathrm{gcm}^{-3}$. This is the characteristic density that defines a Universe with Euclidean geometry, $K = 0$.
We can now define the *density parameter*, which is useful to describe the amount of density, and, therefore, energy that is contributed by each component that composes the Universe today and with time dependence:

$$\Omega_s \equiv \frac{\rho_s(t)}{\rho_{crit}(t)}. \tag{1.14}$$

We use s to indicate the different components of the Universe.

[1] for simplicity the time dependence of $a(t)$ is made implicit

parameter	value 68% limits
$\Omega_{0,bar}h^2$	0.0224 ± 0.0001
$\Omega_{0,DM}h^2$	0.120 ± 0.001
$H_0[\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}]$	67.4 ± 0.5
$t_0[\mathrm{Gyr}]$	13.801 ± 0.024
$\Omega_{0,m}$	0.3111 ± 0.0056
$\Omega_{0,\Lambda}$	0.6889 ± 0.0056

Table 1.1 *Cosmological parameters from Planck 2018 (Aghanim et al., 2020). Here we report the density parameters for baryons and dark matter, the Hubble constant, the age of the Universe, the matter density parameter, and the cosmological constant density parameter.*

1.1.1 The ΛCDM model

The currently favored cosmological model often refers to as the standard or the *concordance* model, is known as the ΛCDM model. We are going to describe the elements of the model. As a reference, we take the values of the main cosmological parameters, derived from the most recent analysis of the Cosmic Microwave Background Radiation data of the *Planck mission* (Aghanim et al., 2020) in table 1.1.

The model considers three main contributes to the energy, described as "cosmic fluids" by Eq. 1.11:

- **Matter** or *dust*: they are massive non-relativistic particles at rest in the comoving frame; the pressure of matter is thus negligible, from the adiabatic Eq. 1.11 $\rho \propto a^{-3}$. This trend means that, as expected, density decreases only because the volume increases.

- **Radiation**: this fluid is made up of photons or, more in general, relativistic particles. In the models in which neutrinos are massless, they fall into this category. We would have $\rho \propto a^{-4}$ for radiation. Photons also lose energy because of the cosmological redshift, and thus density decreases more steeply than matter.

- **Dark energy**: different from any other ordinary cosmic fluid that can only decelerate the expansion of the Universe. The cosmological constant Λ is the most basic form of dark energy and, in the standard model, is the preferred one.

As said before, dark energy is introduced to justify the accelerated expansion proved by distant supernovae observations. It is the cosmological constant that represents dark energy (Λ in ΛCDM) in the concordance model; we have then the

density parameter:

$$\Omega_\Lambda = \frac{\Lambda c^2}{3H^2(t)}, \tag{1.15}$$

and dominates the content of the Universe at the current time.

Considering the equations for a flat Universe and using the density parameters of the different components and their scaling with redshift z, the Friedmann Equation reads:

$$E^2(z) \equiv \frac{H^2(z)}{H_0^2} = \left[\Omega_{m,0}(1+z)^3 + \Omega_{\gamma,0}(1+z)^4 + \Omega_{\Lambda,0}\right], \tag{1.16}$$

where it was introduced the *expansion function* as $E(z) \equiv H(z)/H_0$.

Along with dark energy, the main assumption for the standard model is the nature of matter. Today we have evidence of two kinds of matter: ordinary or *baryonic matter*, made up by the particles of the Standard Model of Particle Physics; *dark matter*, made up by unidentified yet particles that interact only gravitationally with each other and the baryons. We have now several observational evidence for the existence of dark matter, for example, the rotation curves of neutral gas in spiral galaxies, the mass estimate of clusters of galaxies thanks to optical, X, and lensing observations, the theory of cosmic structure (see Section 1.2). What is still up to study is the nature of dark matter particles. The basic property has to be the weak interaction with the other content of the Universe; thus, the models differ in two main characteristics: the prediction about the mass of the particle, which defines the time in which dark matter stops to be relativistic, and the interaction with radiation and other particles (*decoupling time*). The standard model predicts a *Cold Dark Matter* (CDM), as opposed to *Hot Dark Matter* (or the intermediate models of *Warm Dark Matter*), in which the particles are extremely weakly interactive and with a high mass therefore non-relativistic. Studies of the large-scale structure of the Universe suggest the cold nature of the dark matter, and this is consistent with a *bottom-up* scenario for cosmic structure formation, where small halos were first to be born and then merged into bigger halos by hierarchical aggregation. The content of matter in the Universe, which produces a density parameter today of $\Omega_{m,0} \approx 0.3$, is $\approx 85\%$ dark matter and $\approx 15\%$ baryonic matter.

As for the relativistic component, it is possible to define two candidates today, photons and neutrinos (if massless as predicted by the standard particle model). The energy density contributed by photons today is determined by the temperature of the Cosmic Microwave Background, black-body radiation with temperature $T_{CMB,0} = 2.73$ K. The density parameter of the radiation is $\Omega_{rad,0} = 2.4 \times 10^{-5} h^{-2}$. The radiation energy content of the Universe today is negligible with respect to the other components.

Like photons, neutrinos were produced in thermal equilibrium in the hot early phase of the Universe. Interacting weakly, they decoupled from the cosmic plasma when the temperature of the Universe was $K_B T \approx 1$ MeV. When the temperature of the Universe dropped to $K_B T \approx 0.5$ MeV, electron-positron pairs annihilated to produce gamma rays, The annihilation heated up the photons but not the neutrinos which had decoupled earlier. Hence the neutrino temperature is lower than the photon temperature by an amount determined by entropy conservation. It can be shown $T_{\nu,0} = \left(\frac{4}{11}\right)^{\frac{1}{3}} T_{CMB,0} = 1.95$ K. The density parameter of one neutrino species (three species in total exist) is $\Omega_{\nu,0} = 2.8 \times 10^{-6} h^{-2}$. On the other hand, if neutrinos have a mass, they should be considered in the density parameter of matter, Ω_m. Some theoretical models suggest that neutrinos might be potential

candidates for the elusive dark matter particle. If this is the case, they would belong to the category of *hot dark matter*, due to their low mass. However, even assuming the highest estimates available for neutrino's mass, they would still contribute only a small proportion to the energy content detected for dark matter.

Measurements from the CMB indicate that the total density parameter of the Universe is consistent with being $\Omega_{tot,0} = 1$. These measures and the evidence for the acceleration of the Universe led to theorizing the existence of a Cosmological constant with $\Omega_{0,\Lambda} \approx 0.7$.

One of the main sources of debate around the standard model that is yet uncertain is the value of the Hubble constant; it should, in general, be $H_0 \approx 70\text{kms}^{-1}\text{Mpc}^{-1}$. Independent measurements of the constant yield different and thus not consistent values arising from the so-called *tension* in the measures of the Hubble constant. Planck analysis yields $H_0 = (67.4 \pm 0.5)\text{kms}^{-1}\text{Mpc}^{-1}$. Local measurements based on the period-luminosity relation of the Cepheid yield $H_0 = (74 \pm 3)\text{kms}^{-1}\text{Mpc}^{-1}$, therefore we have a discrepancy at the $\sim 2\sigma$ level. From gravitational lensing, it is possible to estimate the Hubble constant by measuring the time delay of multiple images: measurements of that kind produce $H_0 = (79 \pm 4)\text{kms}^{-1}\text{Mpc}^{-1}$, for the details Schneider (2006).

1.2 Density perturbations and their evolution

The cosmological principle that states the homogeneity of the matter distribution is actually valid only on a very large scale of the universe. On smaller scales, the observations of galaxies show us that the distribution is not uniform or random. Galaxy positions are correlated and grouped together in galaxy clusters, which in turn have a correlated position and form superclusters. We refer to the three-dimensional distribution of galaxies as the *large-scale structure*.

The standard cosmological model assumes as the origin of the structure formation the small fluctuations in the metric tensor that caused, in turn, fluctuations also in the potential gravitational field and in the density field that take place at some very early initial time. Modern inflationary models suggest that the fluctuations have a quantistic origin. In this case, the fluctuations are uncorrelated, and the distribution of their amplitudes is Gaussian. It is thanks to the gravitational instability that density fluctuations can grow in amplitude. We describe the *density contrast field* as:

$$\delta(\boldsymbol{x}, t) \equiv \frac{\rho(\boldsymbol{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)}, \tag{1.17}$$

where the mean cosmic matter density at time t is $\bar{\rho}(t)$. From today's observations of galaxies, we have $|\delta| \gg 1$ but the study of temperature anisotropies in the CMB suggests instead that at $z \sim 1000$, $|\delta| \ll 1$. We are going to do an overview of the theory of the evolution of density fluctuations using the concepts of the Jeans theory that describes the evolution with linear perturbation theory. This description holds while the density contrast is consistently lower than 1. We can describe the evolution of density fluctuations only if we consider the history of the Universe in its different cosmic epochs. one fundamental timestep is the *equivalence* (a_{eq}) that divides the period in which the density is dominated by radiation from the domination of matter. After the equivalence, the Universe was dominated by weakly interacting dark matter (when $a \gg a_{eq}$). Today it is the cosmological constant, i.e. dark energy, to dominate in the density of the universe.

At different timescales, the perturbations evolve differently according to their nature (radiation, dark matter, or baryonic matter) but can act differently also depending on the (comoving) wavelength of the perturbation itself, a relevant scale length is the *horizon scale*, that is the scale of the causally connected regions in the Universe:

$$\lambda_H(t) = a(t) \int_0^t \frac{c}{a(t')} \mathrm{d}t'. \tag{1.18}$$

Its value depends deeply on the cosmological model and parameters, but in general, it is inversely proportional to the *Hubble parameter* $\lambda_H \propto cH^{-1}$. A perturbation of (comoving) wavelength λ is said to "enter the horizon" when $\lambda = \lambda_H(a)$.
From perturbation theory for linear fluctuations ($\delta \ll 1$) we have that a perturbation which has not entered the horizon at a time with scale factor a ($\lambda > \lambda_H(a)$) grows like:

$$\delta(a) \propto \begin{cases} a^2 & \text{if } a < a_{eq} \\ a & \text{if } a > a_{eq} \end{cases} \tag{1.19}$$

The cosmic microwave background originated at $a \approx 10^{-2} \gg a_{eq}$, thus, the evolution of its fluctuations should go with a. Since the Sachs-Wolfe effect (Wolfe, 1967) states that the temperature fluctuations reflect the density fluctuations δ at the same order of magnitude, the observations of CMB temperature fluctuations of $\frac{\delta T}{T} \approx 10^{-5}$ on large scales means that at $a \approx 10^{-3}$ we had $\delta \approx 10^{-5}$. We should have thus density fluctuations of the level of 10^{-2} today, and the non-linear regime $\delta \sim 1$ should not have been reached. This discrepancy is resolved by considering that the fluctuations studied with CMB observations are relative only to the baryonic matter, which is a small part of all matter content. Dark matter, in fact, as a component that only couples through weak interactions, has fluctuations that could grow as soon as it decoupled from the cosmic fluid. This happened well before the decoupling of the baryons that yields the CMB. Dark matter fluctuations could therefore grow and reach easily the amplitudes observed today. The growth of baryons fluctuations is then sped up by the existing potential of the already evolved dark matter halos: this phenomenon is often referred to as *baryon catch-up* and is one of the strongest pieces of evidence for the existence of dark matter. It is impossible to generalize Eq. 1.19, defined in an Einstein-De Sitter universe (Einstein and De Sitter, 1932) ($\Omega_{m,0} = 1$), to a relation for a curved universe in a simple analytical way. However, numerical simulations show that the effect of curvature is negligible in the early universe and became effective only at $a > a_{eq}$. In an open universe $\Omega_{tot,0} < 1$ and that leads to a gravitational pull of the density fluctuations being weaker and the expansion of the Universe stronger with respect to the flat Universe, then fluctuations at $a > a_{eq}$ evolve as $\delta(a) \propto a^q$ where $q < 1$. On the contrary, if $\Omega_{tot,0} > 1$ and the Universe is closed, the expansion is weaker and the gravitational pull stronger, yielding $\delta(a) \propto a^q$ where $q > 1$ for fluctuations growing after the equivalence.

We understood that for $\lambda > \lambda_H(a)$, the perturbations grow for every a. When they enter the horizon, this is not true in general because gravity is not the only relevant interaction between particles, and other physical interactions must be considered too. If $\lambda < \lambda_H(a_{eq})$, the perturbation enters the horizon while radiation is still dominating the expansion. The *Hubble time* $\tau_H = H^{-1}$, the time-scale of the expansion, is determined by the radiation density ρ_{rad}, and is shorter than the collapse time-scale of the dark matter, as $(G\rho_{rad})^{-1/2} < (G\rho_{DM})^{-1/2}$. Dark matter

perturbations cannot collapse because the expansion driven by the radiation is faster than the collapse of dark matter. The horizon size at the equivalence $\lambda_H(a_{eq})$ is an important scale for structure formation.

We define a_{hor} as the scale factor at which the perturbation with comoving wavelength λ enters the horizon and can be found by solving $\lambda = \lambda_H(a_{hor})$. Then:

$$\lambda \propto a^n \begin{cases} a_{hor} & a_{hor} \ll a_{eq} \\ a_{hor}^{1/2} & a_{eq} \ll a_{hor} \ll a_0. \end{cases} \tag{1.20}$$

It is possible to decompose the density contrast δ into Fourier modes: this is convenient as the Fourier components evolve independently in linear perturbation theory. The Fourier transform of $\delta(\boldsymbol{x})$ is

$$\hat{\delta}(\boldsymbol{k}) = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} \delta(\boldsymbol{x}) e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \mathrm{d}^3\boldsymbol{k} \tag{1.21}$$

Let us note that the dimensions of $\hat{\delta}(\boldsymbol{k})$ are the ones of a volume and that, since $\delta \in \mathbb{R}$ we must have $\hat{\delta}(\boldsymbol{k})^* = \hat{\delta}(-\boldsymbol{k})$. Remembering the definition and effect of the three-dimensional Dirac's delta distribution [2] , we have that the assumed Gaussian density fluctuations $\delta(\boldsymbol{x})$ at the comoving position $\boldsymbol{x}$ can completely be characterized by their *Power Spectrum* $\mathcal{P}(k)$, that we will properly define alongside the other cosmological statistical properties in Section 1.2.1:

$$\langle \hat{\delta}(\boldsymbol{k}) \hat{\delta}^*(\boldsymbol{k'}) \rangle = (2\pi)^3 \mathcal{P}(k) \delta_D^{(3)}(\boldsymbol{k} - \boldsymbol{k'}), \tag{1.22}$$

where the asterisk denotes complex conjugation. The evidence of the curvature of the universe being negligible at early times resolves the fact that the decomposition is possible only in flat space. Moreover, at the late stages of the evolution, the scales $\lambda = \frac{2\pi}{k}$ of the density perturbations are much smaller than the curvature radius of the Universe. The power spectrum is proportional to the average square amplitude of the perturbations with wave vector k, $\mathcal{P}(k \propto \langle |\delta(\boldsymbol{k})|^2 \rangle)$. Averaging in the Fourier space means averaging every wave vector $\boldsymbol{k}$ whose wave number is $|\boldsymbol{k}| = k$. Let us now consider the power spectrum when it was formed at the end of the inflation, i.e. the primordial power spectrum: $\mathcal{P}_i(k) \propto |\delta(\boldsymbol{k})|^2$: if Eq. 1.19 holds, the power spectrum grows as $\mathcal{P}(k; a) \propto a^{2n}$. At the epoch a_{hor}, Eq. 1.20 and the fact that $k \propto \lambda^{-1}$ show that the primordial power spectrum has evolved in:

$$P(k; a_{hor}) \propto a_{hor}^{2n} \mathcal{P}_i(k) \propto \begin{cases} k^{-4} \mathcal{P}_i(k) & a_{hor}(k) < a_{eq} \\ \mathcal{P}_i(k) & a_{hor}(k) > a_{eq}, \end{cases} \tag{1.23}$$

where $a_{hor}(k) < a_{eq}$ describes the scales that enter the horizon before the equivalence (thus with growth suppressed), and $a_{hor}(k) > a_{eq}$ are the large scales (small k) not reached by the horizon before the equivalence.

[2] 3-dimensional Dirac's delta:

$$\delta_D^{(3)}(\boldsymbol{k}) = \frac{1}{(2\pi)^3} \int_{\mathcal{R}^3} e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \mathrm{d}^3\boldsymbol{x},$$

$$f(\boldsymbol{x}) = \int_{\mathcal{R}^3} f(\boldsymbol{k}) \delta_D^{(3)}(\boldsymbol{x} - \boldsymbol{k}) \mathrm{d}^3\boldsymbol{k}.$$

We can assume that the total power of the density fluctuations at a_{hor} should be scale-invariant, then $k^3\mathcal{P}(k) = \text{const.}$ or $\mathcal{P}(k) \propto k^{-3}$. This leads to the primordial *Harrison-Zel'dovich spectrum* $\mathcal{P}_i(k) \propto k$ (Harrison (1970), Zeldovich (1972)). Thus Eq. 1.23 can be written as:

$$\mathcal{P}(k; a_{hor}) \propto \begin{cases} k^{-3} & a_{hor}(k) < a_{eq} \\ k & a_{hor}(k) > a_{eq}, \end{cases} \tag{1.24}$$

Different dark matter models that we introduced in 1.1.1 can be studied also in relation to the evolved primordial power spectrum. When dark matter particles have a velocity comparable to the speed of light, if the density perturbation does not reach a certain minimum mass or a minimum size, the perturbation can be damped away by these free streaming particles. At large k (small λ) the density perturbation spectrum has an exponential cut-off. We have then models of *hot dark matter* whose particles are light and damp away perturbations at small-scale perturbations, while the *cold dark matter* models predict massive particles slow enough to cause no significant damping.

1.2.1 Statistical tools for Cosmology

We introduced the power spectrum of the fluctuations to study the evolution of their growth, which is a powerful tool that is defined in the context of the statistical approach to Cosmology. We are now describing some essential statistical elements in order to explore more in the detail the methods and the application of the cosmological statistical analysis of the density contrast field $\delta(\boldsymbol{x})$.

Let us consider a random field $g(\boldsymbol{x})$ whose expectation value is zero everywhere. In the case of a non-null expectation value, we could consider the difference between the field and the expectation value in each point of the space. The field could be both real and complex and defined in a n-dimensional space of the coordinates $\boldsymbol{x}$. Moreover, the field can be:

- *homogeneous* if $g(\boldsymbol{x})$ is statistically indistiguishible from $g(\boldsymbol{x} + \boldsymbol{r})$, where $\boldsymbol{r}$ is an arbitrary translation vector;
- *isotropic* if $g(\boldsymbol{x})$ is statistically indistiguishible from $g(\mathcal{R}\boldsymbol{x})$, where $\mathcal{R}$ is an arbitrary rotation matrix in n dimensions.

It is possible to describe the *level of order* of the field introducing the *correlation function*, which depicts how variables are connected to each other and measures how the field co-varies in average over space:

$$\xi(r) = C_{gg}(|\boldsymbol{x} - \boldsymbol{x'}|) = \langle g(\boldsymbol{x}) g^*(\boldsymbol{x'}) \rangle, \tag{1.25}$$

which for a homogeneous field depends only on the absolute value of the difference vector between the two points $\boldsymbol{x}$ and $\boldsymbol{x'} = \boldsymbol{x} + \boldsymbol{r}$.
The correlation function is actually averaged twice. First, we fix $\boldsymbol{x}$ and compute $g(\boldsymbol{x})g^*(\boldsymbol{x'})$ for every $\boldsymbol{x'} = \boldsymbol{x} + \boldsymbol{r}$ that is distant $r = |\boldsymbol{r}|$ from $\boldsymbol{x}$ and average all these values. Then, we repeat for every $\boldsymbol{x}$ and average again. When g is the density contrast $\delta(\boldsymbol{x})$ defined in 1.17 the correlation function is the observable used in *clustering studies*.
The correlation function is real, even if the field is complex: we can prove that by taking the complex conjugate of Eq. 1.25, which is equivalent to interchanging $\boldsymbol{x}$

and $\boldsymbol{x'}$, leaving the right-hand-side unaffected. Let us define the Fourier-transform pair of g:

$$\hat{g}(\boldsymbol{k}) = \int_{\mathbb{R}^n} g(\boldsymbol{x}) e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \mathrm{d}^n\boldsymbol{x}. \tag{1.26}$$

$$g(\boldsymbol{k}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{g}(\boldsymbol{k}) e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \mathrm{d}^n\boldsymbol{k}. \tag{1.27}$$

The correlation function in the Fourier space is then:

$$\langle \hat{g}(\boldsymbol{k})\hat{g}^*(\boldsymbol{k'})\rangle = \int_{\mathbb{R}^n} e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \int_{\mathbb{R}^n} \langle g(\boldsymbol{x}) g^*(\boldsymbol{x'})\rangle e^{i\mathbf{k}'\cdot\boldsymbol{x'}} \mathrm{d}^n\boldsymbol{x}\mathrm{d}^n\boldsymbol{x'}. \tag{1.28}$$

Using Eq. 1.25 and the fat that $\boldsymbol{x'} = \boldsymbol{x} + \boldsymbol{r}$, this becomes:

$$\begin{aligned}\langle \hat{g}(\boldsymbol{k})\hat{g}^*(\boldsymbol{k'})\rangle &= \int_{\mathbb{R}^n} e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \int_{\mathbb{R}^n} C_{gg}(|\boldsymbol{r}|) e^{i\mathbf{k}'\cdot\boldsymbol{x}+\boldsymbol{r}} \mathrm{d}^n\boldsymbol{x}\mathrm{d}^n\boldsymbol{x'} = \\ &= (2\pi)^n \delta_D^{(n)}(\boldsymbol{k}-\boldsymbol{k'}) \int_{\mathbb{R}^n} C_{gg}(|\boldsymbol{r}|) e^{-i\mathbf{k}\cdot\boldsymbol{r}} \mathrm{d}^n\boldsymbol{r} = \\ &= (2\pi)^n \delta_D^{(n)}(\boldsymbol{k}-\boldsymbol{k'}) \mathcal{P}_g(k).\end{aligned} \tag{1.29}$$

In the final passage, we defined the power spectrum of the isotropic random field g as

$$\mathcal{P}_g(k) = \int_{\mathbb{R}^n} C_{gg}(|\boldsymbol{r}|) e^{-i\mathbf{k}\cdot\boldsymbol{r}} \mathrm{d}^n\boldsymbol{r}, \tag{1.30}$$

thus we have that the power spectrum is the Fourier transform of the two-point correlation function. It is isotropy that allows $\mathcal{P}_g$ to depend only on the modulus of $\boldsymbol{k}$.

Gaussian random fields are characterized by the property that the probability distribution of any linear combination for the random field $g(\boldsymbol{x})$ is Gaussian. Thus the joint probability distribution of a number M of linear combinations of the random variable $g(\boldsymbol{x_i})$ is a multivariate Gaussian. That is equivalent to require that the Fourier transform components $\hat{g}(\boldsymbol{k})$ are mutually statistically independent and that the probability densities for the $\hat{g}(\boldsymbol{k})$ are Gaussian with dispersion $\mathcal{P}_g(k)$. Thus, a Gaussian random field is fully characterized by its power spectrum.

1.2.2 The cosmological power spectrum

The statistical approach used in Cosmology and clustering studies relies on the so-called *ergodic hypothesis*: the mean value of a given quantity, when averaged over a large number of volumes (each representing a realization of the Universe), is equal to the average of the mean values of the quantity averaged on many sub-volumes of a single realization of the Universe. For the density contrast probability distribution, the ergodic hypothesis is always verified as it is Gaussian. It is necessary to consider sub-volumes that constitute a fair sample and are large, independent, and a good representation of the Universe (isotropic). Let us now consider the actual expression for the δ probability distribution and apply the tools we described in Section 1.2.1. As a Gaussian distribution with zero mean, $P(\delta)$ is described only by its variance σ^2:

$$P(\delta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\delta^2}{2\sigma^2}} \tag{1.31}$$

This is an approximation since for $\delta < -1$, the distribution must be truncated. From Eq. 1.30 and 1.29 we have the correlation function of density constrast field $\delta(\boldsymbol{x})$ can be expressed as:

$$\xi(r) = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} \mathcal{P}(k) e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \mathrm{d}^3\boldsymbol{k}, \tag{1.32}$$

where in turn we can write the power spectrum as:

$$\langle \hat{\delta}(\boldsymbol{k}) \hat{\delta}^*(\boldsymbol{k'}) \rangle = (2\pi)^3 \delta_D^{(n)}(\boldsymbol{k} - \boldsymbol{k'}) \mathcal{P}(k), \tag{1.33}$$

and it has the dimensions of a volume. It is important to notice that $\mathcal{P}(k)$ is a power density while the actual "power" is $\mathcal{P}\mathrm{d}^3\boldsymbol{k}$, i.e. the amplitude of the mode k of the correlation function. We found that the power spectrum is proportional to the average square amplitude of a perturbation with wave vector k:

$$\mathcal{P}(k) \propto \langle \hat{\delta}(\boldsymbol{k}) \hat{\delta}^*(\boldsymbol{k'}) \rangle \propto \langle |\hat{\delta}(\boldsymbol{k})|^2 \rangle, \tag{1.34}$$

remembering Dirac's delta properties. Expanding the Dirac's delta, we find the proportionality constant of the power spectrum representing the *volume of the Universe*, i.e. an integral over all $\mathbb{R}^3$, and thus we have:

$$\mathcal{P}(k) = \frac{\langle |\hat{\delta}(\boldsymbol{k})|^2 \rangle}{V_\infty} \tag{1.35}$$

The power spectrum is extremely important as it can give us the only parameter of the δ probability distribution, the variance σ^2. It can be verified that the variance is the second moment of the power spectrum:

$$\sigma^2 = \frac{1}{2\pi^2} \int_0^\infty \mathcal{P}(k) k^2 \mathrm{d}k. \tag{1.36}$$

As we discussed in Section 1.2, the initial power spectrum given by the inflationary expansion phase of the Universe is a power law that has no privileged scale:

$$\mathcal{P}(k) = Ak^n. \tag{1.37}$$

The amplitude must be constrained with the observations while the inflationary model suggests that $n = 1$ is in the primordial power spectrum, defining the scale-invariant *Zel'dovich Spectrum*. The evolution of the perturbations modifies the power spectrum in a scale-dependent way so that it can usually be described as a local power law with local index $n_{eff}(k)$:

$$n_{eff}(k) = \frac{\mathrm{d} \ln \mathcal{P}(k)}{\mathrm{d}k}. \tag{1.38}$$

From Eq. 1.23, the time evolution of the power spectrum is given by:

$$\mathcal{P}(k, t) = \mathcal{P}_i(k) \delta_+^2(t), \tag{1.39}$$

where δ_+ is equivalent to a_{hor}^q in Eq. 1.23.

It is very useful to see the variance of the distribution as a function of the mass of the structures. In order to do this, we study the time evolution of the variance itself:

$$\sigma^2 \propto \delta_+^2(t) k^{n+3} \propto \delta_+^2(t) \lambda^{-(n+3)} \propto \delta_+^2(t) M^{-\frac{n+3}{3}} \tag{1.40}$$

where we used the power law spectrum of Eq. 1.37 in the definition of the variance 1.36, and the proportionality relation between $M \propto \lambda^3 \propto k^{-3}$.

The Zel'dovich spectrum can be obtained by considering that the inflation causes perturbations in the spacetime metric that become perturbations in the gravitational potential that have no privileged scale and have a constant mean amplitude (white noise). For a fixed R the fluctuations of the potential are:

$$\begin{aligned} \delta\Phi &\propto \frac{G\delta M}{R} \propto \frac{G\delta\rho R^3}{R} \propto \delta\rho R^2 \propto \sigma R^2 \\ \delta\Phi &\propto \sigma M^{\frac{2}{3}} \propto M^{-\frac{n+3}{6}+\frac{2}{3}} \propto M^{\frac{1-n}{6}} \end{aligned} \tag{1.41}$$

To have a constant amplitude, the power law index must be $n = 1$.

Chapter 2

Gravitational lensing

This Chapterwill describe the fundamental aspects of the theory of *Gravitational lensing*. In the context of this Thesis, we will pay particular attention to the weak-lensing regime, as opposed to the strong-lensing regime, in which the strength of the gravitational deflection of light is reduced, and is possible to use the mathematical background at its first-order approximation. This introduction is divided into two sections: Section 2.1 will revise some theoretical basics elements of the phenomenon, from the lens equation to the E/B decomposition that is crucial to this work; while in Section 2.2 we focus on the tools that are used in weak-lensing studies and in particular on the statistics known as the *Aperture mass*. The main references for this Chapter are Bartelmann and Schneider (2001) and Umetsu (2020).

The local universe appears to be highly inhomogeneous on a wide range of scales, from galaxies, through galaxy groups and clusters, to forming superclusters, large-scale filaments, and cosmic voids. Light rays from far background sources are deflected when propagating through the gravitational field caused by such local inhomogeneities along the line of sight. This phenomenon, the deflection of light, was theoretically predicted by Einstein in the theory of General Relativity. Still, several scientists already in the past had speculated about it, the most illustrious example being sir Isaac Newton who deducted it in 1704. However, with Newtonian gravitational physics, the computation of the deflection predicts an effect that is half of the real effect computed with General Relativity. This was confirmed thanks to a mission guided by Eddington having the goal of measuring the deflection of light grazing the surface of the Sun during a solar eclipse (Dyson et al., 1920): from Newton's theory, the deflection would have been of 0.87 arcsecs while the value predicted by General Relativity and confirmed by the observations is 1.75 arcsec.
Even though, for Einstein himself, the deflection of light at the rim of the Sun marked a confirmation of his theory of General Relativity, he was sure that this effect would hardly ever gain astrophysical relevance. Fritz Zwicky in 1937 proposed the idea that clusters of galaxies could act as deflectors of light or *gravitational lenses* after discovering the first indication of dark matter from the observation of the Coma cluster (Zwicky, 1933).
Nonetheless, it took until 1979 for the first gravitational lens to be found: it was discovered the first multiple image system, an effect peculiar of the strong lensing regime (there may exist more than one null geodesic connecting the world-line of a source with the observation event). Alongside the possibility of multiplying the

images of background sources, gravitational lensing does not change the surface brightness of a source, but it changes the effective solid angle under which its light can be seen and increases its flux: this is the *magnification*, the property that allows the use of gravitational lenses as "cosmic telescopes". The lensing effect on which this Thesis is based is the deformation of the images of resolved sources. This deformation can either be strong, as for gravitational arcs and Einstein rings or weak, like a change in the source's ellipticity. Although weak distortions in individual images can hardly be recognized, and the net distortion averaged over an ensemble of images can still be detected. This last possibility marks out the *weak lensing* regime, which does not have evident effects, and all the measurements of its effect are statistical in nature. Weak lensing measurements are based on the observation of the galaxy shape, i.e. the ellipticity, in wide field surveys in order to constrain the distortion pattern given by the lensing of generally large distribution of matter in an intermediate position between the galaxies in the background and the observer.

Magnification and distortion effects due to weak lensing are used to probe the statistical properties of the matter distribution between us and an ensemble of distant sources. Moreover, since the deflection angle is determined only by the gravitational field of the matter distribution along its path, thus light deflection probes the total matter density. This led to the possibility of directly investigating the distribution of dark matter and influenced our understanding of structure formation, supporting hierarchical structure formation in cold dark matter (CDM) models.

2.1 Lensing theory

The deflection angle of a ray of light in position $\boldsymbol{\xi'}$ in the presence of a point-like mass M is (Bartelmann and Schneider, 2001) can be written as:

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4GM}{c^2}\frac{\boldsymbol{\xi} - \boldsymbol{\xi'}}{|\boldsymbol{\xi} - \boldsymbol{\xi'}|^2}. \tag{2.1}$$

We consider $\boldsymbol{\xi}$ the 2-dimensional position on a plane perpendicular to the line of sight. If we consider instead a mass distribution characterized by a surface density on the sky plane $\Sigma(\boldsymbol{\xi'})$, then the deflection angle of a light ray passing in $\boldsymbol{\xi}$ in the plane is

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4G}{c^2}\int_{\mathbb{R}^2} \Sigma(\boldsymbol{\xi'})\frac{\boldsymbol{\xi} - \boldsymbol{\xi'}}{|\boldsymbol{\xi} - \boldsymbol{\xi'}|^2}\mathrm{d}^2\boldsymbol{\xi'}. \tag{2.2}$$

This expression holds as long as the *thin screen approximation* (Meneghetti, 2021) holds: even considering the case of lensing by galaxy clusters, the lens physical size is generally much smaller than the distances between observer, lens and source, thus we can approximate the lens by a planar distribution, the lens plane. A relevant exception, in particular for this work, is the deflection by the large-scale structures.

The basic tool for the gravitational lensing theory is the *lens equation* that relates the true position of the source to its observed position in the sky. In Figure 2.1, the source and lens are identified by planes perpendicular to a straight line (*optical axis*) from the observer to the lens at the distance of the source and of the lens, respectively, let $\boldsymbol{\eta}$ be the 2-dimensional position of the source on its plane. We can obtain geometrically:

$$\boldsymbol{\eta} = \frac{D_S}{D_L}\boldsymbol{\xi} - D_{LS}\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) \tag{2.3}$$

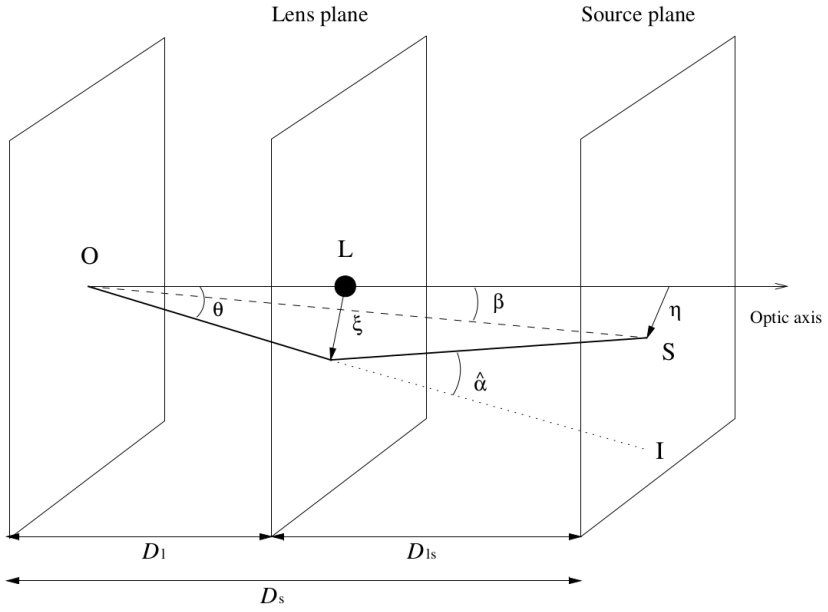


Figure 2.1 *Illustration of a typical lens system. The light ray propagates from the source (S) at the position $\boldsymbol{\eta}$ in the source plane to the observer (O), passing the position $\boldsymbol{\xi}$ in the lens plane (L), resulting in a bending angle $\hat{\boldsymbol{\alpha}}$. The angular position of the source (S) relative to the optical axis is denoted by $\boldsymbol{\beta}$, and that of the image (I) relative to the optical axis is denoted by $\boldsymbol{\theta}$. The quantities D_L, D_S, and D_{LS} are the observer-lens, observer-source, and lens-source angular diameter distances, respectively (Umetsu, 2020).*

being D_S, D_L and D_{LS} the angular diameter distances between observer and source, between observer and lens and between lens and source, in general, $D_S \neq D_L + D_{LS}$, and the equality holds only in the local Universe, where we can ignore the expansion of the Universe and the geometry is Euclidean. Introducing angular coordinates $\boldsymbol{\eta} = D_S\boldsymbol{\beta}$ and $\boldsymbol{\xi} = D_L\boldsymbol{\theta}$ Eq. 2.3 becomes

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \frac{D_{LS}}{D_S}\hat{\boldsymbol{\alpha}}(D_L\boldsymbol{\theta}) \equiv \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}), \tag{2.4}$$

where we defined the *reduced deflection angle* in the last step.

A useful lensing quantity is the *lensing potential* of the lens, defined as a scaled integral of the gravitational potential ψ of the lens on the *line of sight* (l.o.s.):

$$\phi(\boldsymbol{\theta}) \equiv \frac{D_{LS}}{D_L D_L}\frac{2}{c^2}\int_{\mathbb{R}} \psi(D_L\boldsymbol{\theta}, x_{los})\mathrm{d}x_{los}, \tag{2.5}$$

where we consider the three-dimensional position of the lens $\boldsymbol{\xi} + \boldsymbol{x_{los}}$ is the sum of a vector $\boldsymbol{\xi} = D_L\boldsymbol{\theta}$ on the lens plane plus a vector $\boldsymbol{x_{los}}$ on the l.o.s. .

The lensing potential is adimensional. A relation between reduced deflection angle and lensing potential can be found:

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\psi(\boldsymbol{\theta}). \tag{2.6}$$

Then if we define the *critical surface density* of a lensing system

$$\Sigma_{crit} \equiv \frac{c^2}{4\pi G}\frac{D_S}{D_{LS}D_L}, \tag{2.7}$$

it is possible to define the *convergence* κ, an adimensional surface density:

$$\kappa(\boldsymbol{\theta}) \equiv \frac{\Sigma(\boldsymbol{\theta})}{\Sigma_{crit}}. \tag{2.8}$$

The Poisson equation projected on the lens plane is then:

$$\nabla^2_{\boldsymbol{\theta}}\psi(\boldsymbol{\theta}) = 2\kappa(\boldsymbol{\theta}). \tag{2.9}$$

It is possible to express the lensing potential as a convolution if we express the deflection angle in Eq. 2.2 as:

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{1}{\pi}\int_{\mathbb{R}^2}\kappa(\boldsymbol{\theta'})\frac{\boldsymbol{\theta}-\boldsymbol{\theta'}}{|\boldsymbol{\theta}-\boldsymbol{\theta'}|^2}\mathrm{d}^2\boldsymbol{\theta'}; \tag{2.10}$$

then using it in Eq. 2.6 we have:

$$\psi(\boldsymbol{\theta}) = \frac{1}{\pi}\int_{\mathbb{R}^2}\kappa(\boldsymbol{\theta'})\ln|\boldsymbol{\theta}-\boldsymbol{\theta'}|\mathrm{d}^2\boldsymbol{\theta'}. \tag{2.11}$$

2.1.1 First-order lens mapping and lensing regimes

The solutions $\boldsymbol{\theta}$ of the lens equation yield the angular positions of the images of a source at $\boldsymbol{\beta}$. The shapes of the images will differ from the intrinsic shape of the source because light bundles are deflected differentially. In general, the shape of the images must be determined by solving the lens equation for all points within an extended source. Liouville's theorem and the absence of emission and absorption of photons in gravitational light deflection imply that lensing conserves the surface brightness (or specific intensity). Hence, if $I^{(s)}(\boldsymbol{\beta})$ is the surface brightness distribution in the plane of the source, the observed surface brightness distribution in the lens plane would be

$$I(\boldsymbol{\theta}) = I^{(s)}[\boldsymbol{\beta}(\boldsymbol{\theta})]. \tag{2.12}$$

If a source is much smaller than the angular scale on which the lens properties change, the lens mapping can be linearised. The lens equation in the first-order lens mapping is:

$$\boldsymbol{\beta} = \mathcal{A}(\boldsymbol{\theta})\boldsymbol{\theta}, \tag{2.13}$$

the distortion of the images is described by the Jacobian matrix

$$\mathcal{A}(\boldsymbol{\theta}) = \frac{\partial\boldsymbol{\beta}}{\partial\boldsymbol{\theta}} = (\delta_{ij} - \psi_{ij}(\boldsymbol{\theta})) = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}, \tag{2.14}$$

where $\psi_{ij}(\boldsymbol{\theta}) = \frac{\partial^2\psi(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}$.
To understand the last passage, we need to remember that the lensing Jacobian is a symmetric tensor, thus it can be decomposed as the sum of an isotropic part $\mathcal{A}^{iso}_{ij}(\boldsymbol{\theta})$ and an anisotropic one $\mathcal{A}^{aniso}_{ij}(\boldsymbol{\theta})$. The first one is found using the Poisson Eq. 2.9:

$$\mathcal{A}^{iso}_{ij}(\boldsymbol{\theta}) = \frac{1}{2}\mathrm{Tr}\mathcal{A}(\boldsymbol{\theta})\delta_{ij} = \frac{1}{2}[2-(\psi_{11}+\psi_{22})]\delta_{ij} = [1-\kappa(\boldsymbol{\theta})]\delta_{ij}, \tag{2.15}$$

where the convergence is :

$$\kappa(\boldsymbol{\theta}) = \frac{1}{2}[\psi_{11}(\boldsymbol{\theta}) + \psi_{22}(\boldsymbol{\theta})] \,. \tag{2.16}$$

Therefore the isotropic deflection is just a rescale of the vectors in the source plane. The anisotropic component is defined as
$\mathcal{A}_{ij}^{aniso}(\boldsymbol{\theta}) = \mathcal{A}_{ij} - \frac{1}{2}\mathrm{Tr}\mathcal{A}(\boldsymbol{\theta})\delta_{ij}$ and if we take the opposite we obtain the *shear tensor* Γ:

$$\Gamma_{ij}(\boldsymbol{\theta}) = -\mathcal{A}_{ij}^{aniso}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{2}(\psi_{11} - \psi_{22}) & \psi_{12} \\ \psi_{12} & -\frac{1}{2}(\psi_{11} - \psi_{22}) \end{bmatrix} = \begin{bmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{bmatrix}. \tag{2.17}$$

It is a symmetric tensor with null trace with two independent *shear components*:

$$\gamma_1(\boldsymbol{\theta}) = \frac{1}{2}[\psi_{11}(\boldsymbol{\theta}) - \psi_{22}(\boldsymbol{\theta})], \tag{2.18}$$

$$\gamma_2(\boldsymbol{\theta}) = \psi_{12}(\boldsymbol{\theta}). \tag{2.19}$$

We can define the *shear* as the complex number $\gamma = \gamma_1 + i\gamma_2$, its *shear* modulus $|\gamma| = \sqrt{\gamma_1^2 + \gamma_2^2}$, then $\det\Gamma = -|\gamma|^2$ and the eigenvalues of the shear tensor are $\pm|\gamma|$. We finally get Eq. 2.14.

If we consider a point $\boldsymbol{\theta}_0$ within an image and the corresponding point $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(\boldsymbol{\theta}_0)$ inside the source, we can use the locally linearised lens equation to write Eq. 2.12 as:

$$I(\boldsymbol{\theta}) = I^{(s)}\left[\boldsymbol{\beta}_0 + \mathcal{A}(\boldsymbol{\theta}_0)\cdot(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right]. \tag{2.20}$$

According to this equation, the images of a circular source become ellipses. The ratios of the semi-axes of such an ellipse to the radius of the source are given by the inverse of the eigenvalues of $\mathcal{A}(\boldsymbol{\theta}_0)$, which are $1 - \kappa \pm |\gamma|$, and the ratio of the solid angles subtended by an image and the unlensed source is the inverse of the determinant of $\mathcal{A}$: this is possible because the jacobian is a symmetric tensor and can be diagonalized. The fluxes observed from the image and from the unlensed source are given as integrals over the brightness distributions $I(\boldsymbol{\theta})$ and $I^{(s)}(\boldsymbol{\beta})$, respectively, and their ratio defines the *magnification* $\mu(\boldsymbol{\theta}_0)$. From Eq. 2.20, we find:

$$\mu(\boldsymbol{\theta}) = \frac{1}{\det\mathcal{A}(\boldsymbol{\theta})} = \frac{1}{[1 - \kappa(\boldsymbol{\theta}) - |\gamma|(\boldsymbol{\theta})][1 - \kappa(\boldsymbol{\theta}) + |\gamma|(\boldsymbol{\theta})]}. \tag{2.21}$$

The images are thus distorted in shape and size. The shape distortion is due to the tidal gravitational field, described by the shear γ, whereas the magnification is caused by both isotropic focusings caused by the local matter density κ and anisotropic focusing caused by the shear. in Figure 2.2, it is shown the combined effect of shear and convergence.

A lens system that has a region with $\kappa(\boldsymbol{\theta}) > 1$ can produce multiple images for certain source positions β, and such a system is referred to as being *supercritical* because the surface density is greater than the critical density. Being supercritical is a sufficient but not necessary condition for a general lens to produce multiple images because the shear can also contribute to multiple imaging. Nevertheless, this provides a simple criterion to broadly distinguish the regimes of multiple and single imaging. Keeping this in mind, we refer to the region where $\kappa(\boldsymbol{\theta}) \gtrsim 1$ as the *strong-lensing regime* and the region where $\kappa(\boldsymbol{\theta}) \ll 1$ as the *weak-lensing regime*.

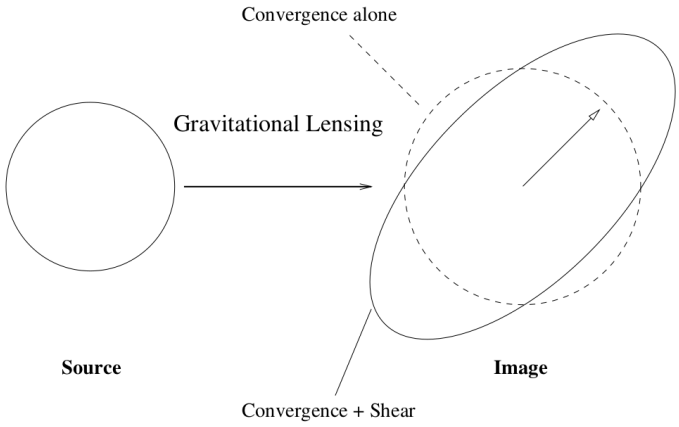


Figure 2.2 *Illustration of the effects of the convergence κ and the shear γ on the angular shape and size of a hypothetical circular source. The convergence acting alone causes an isotropic focusing (magnification) of the image (dashed circle), while the shear deforms it to an ellipse (Umetsu, 2020).*

2.1.2 Properties of the shear tensor

In this Section, we are going into the details of the shear tensor defined in Eq. 2.17. Γ is a symmetric and trace-less tensor and can be expressed in terms of the Pauli matrices as

$$\Gamma(\boldsymbol{\theta}) = \gamma_1(\boldsymbol{\theta})\sigma_3 + \gamma_2(\boldsymbol{\theta})\sigma_1 \,. \tag{2.22}$$

It can also be expressed using the index notation as:

$$\Gamma_{ij}(\boldsymbol{\theta}) = \left(\partial_i\partial_j - \delta_{ij}\frac{1}{2}\nabla^2\right)\psi(\boldsymbol{\theta}), \tag{2.23}$$

where $\partial_i = \frac{\partial}{\partial\theta_i}$, $\nabla^2 = \partial_i^2 + \partial_j^2$, δ_{ij} is Kronecker's delta.

If a rotation $R(\phi)$ is applied to the coordinate system, it is possible to show that the isotropic part of the matrix does not change, and the shear tensor is modified according to $\Gamma' = R^T(\phi)R(\phi)$ and the shear components change as follows:

$$\begin{pmatrix}\gamma_1' \\ \gamma_2'\end{pmatrix} = R^T(2\phi)\begin{pmatrix}\gamma_1 \\ \gamma_2\end{pmatrix}, \tag{2.24}$$

or in an explicit way:

$$\begin{cases}\gamma_1' = +\gamma_1\cos 2\phi + \gamma_2\sin 2\phi \\ \gamma_2' = -\gamma_1\sin 2\phi + \gamma_2\cos 2\phi\end{cases} . \tag{2.25}$$

The shear components are mapped on themselves after a π rotation, so the shear is not a vector, but a spin-2 field (a quantity has spin N if has the same value after a rotation by $\frac{2\pi}{N}$). Therefore the Jacobian matrix can be diagonalized with a rotation that makes $\gamma_1' = \gamma$ and $\gamma_2' = 0$, which can be obtained if the rotation angle is $\phi = \frac{1}{2}\arctan\frac{\gamma_1}{\gamma_2}$.

For the *weak-lensing regime*, it is useful to study the first-order mapping of a circular source centered in $\boldsymbol{\beta} = 0$ with radius β. This means that $\beta^2 = \beta_1^2 + \beta_2^2$. Let

us also suppose that our coordinate axes are the eigenvectors. From Eq. 2.13 we have that:

$$\begin{aligned}\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix} &= \begin{pmatrix}1-\kappa-|\gamma| & 0\\ 0 & 1-\kappa+|\gamma|\end{pmatrix}\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix} = \\ &= \begin{pmatrix}(1-\kappa-|\gamma|)\theta_1 & 0\\ 0 & (1-\kappa+|\gamma|)\theta_2\end{pmatrix}.\end{aligned} \tag{2.26}$$

This means that the circularity relation becomes:

$$\beta^2 = \beta_1^2 + \beta_2^2 = (1-\kappa-|\gamma|)^2\theta_1^2 + (1-\kappa+|\gamma|)^2\theta_2^2 : \tag{2.27}$$

that is the equation of an ellipse; therefore, the circular source is mapped in an elliptical map whose axes are directed along the eigenvectors, and the lengths of the semiaxes are different because of the shear. If a and b are the semiaxis lengths, the ellipticity e of the ellipse is:

$$e = \frac{a+b}{a-b} = \frac{\frac{\beta}{1-\kappa-|\gamma|} + \frac{\beta}{1-\kappa+|\gamma|}}{\frac{\beta}{1-\kappa-|\gamma|} - \frac{\beta}{1-\kappa+|\gamma|}} = \frac{1-\kappa+|\gamma|-1+\kappa+|\gamma|}{1-\kappa+|\gamma|+1-\kappa-|\gamma|} = \frac{|\gamma|}{1-\kappa}. \tag{2.28}$$

The ellipticity of the source is defined *reduced shear modulus*:

$$|g| = \frac{|\gamma|}{1-\kappa}. \tag{2.29}$$

In general, we can define the *Reduced shear tensor* as a scaled version of the shear tensor:

$$\boldsymbol{g}(\boldsymbol{\theta}) = \frac{1}{1-\kappa(\boldsymbol{\theta})}\Gamma(\boldsymbol{\theta}), \tag{2.30}$$

and by analogy with the shear tensor, we can define the reduced shear components g_1 and g_2 and thus the complex reduced shear $g = g_1 + ig_2$. This quantity is crucial to the weak-lensing studies because, the observed ellipticity of an intrinsic circular source, is a direct measure of the reduced shear and this is an effect of the gravitational lensing only. However real galaxies are intrinsically elliptical and we will see how to deal with that. Moreover, in the weak lensing limit where $\kappa \ll 1$ we have that $g \simeq \gamma$, and the reduced shear corresponds to the actual shear.

If we consider the expression for the lensing potential as a convolution, as in Eq. 2.11, it is possible to show that also the complex shear can be written as a convolution of the convergence, as:

$$\gamma(\boldsymbol{\theta}) = \frac{1}{\pi}\int_{\mathbb{R}^2} \kappa(\boldsymbol{\theta'})\mathcal{D}(\boldsymbol{\theta}-\boldsymbol{\theta'})\mathrm{d}^2\boldsymbol{\theta'}, \tag{2.31}$$

with

$$\mathcal{D}(\boldsymbol{\theta}) = \frac{\theta_2^2 - \theta_1^2 - 2i\theta_1\theta_2}{|\boldsymbol{\theta}|^4} = -\frac{1}{(\theta_1 - i\theta_2)^2}. \tag{2.32}$$

and the complex conjugate is:

$$\mathcal{D}^*(\boldsymbol{\theta}) = \frac{\theta_2^2 - \theta_1^2 + 2i\theta_1\theta_2}{|\boldsymbol{\theta}|^4} = -\frac{1}{(\theta_1 + i\theta_2)^2}. \tag{2.33}$$

In addition to this relation between shear and convergence, we can find another one by considering that they are both linear combinations of the second derivatives of the potential. We introduce the notation of the spin operators:

$$\partial = \partial_1 + i\partial_2 \qquad \textit{spin rising operator} \tag{2.34}$$

$$\partial^* = \partial_1 - i\partial_2 \qquad \textit{spin lowering operator}, \tag{2.35}$$

then we have $\nabla^2 = \partial_1^2 + \partial_2^2 = \partial\partial^*$. Finally from Eq. 2.16, Eq. 2.18 and Eq. 2.19 some algebraic calculations show that

$$\kappa = \frac{1}{2}\partial\partial^*\psi(\boldsymbol{\theta}) \tag{2.36}$$

$$\gamma = \frac{1}{2}\partial\partial\psi(\boldsymbol{\theta}) \tag{2.37}$$

$$\nabla^2\kappa(\boldsymbol{\theta}) = \partial^*\partial^*\gamma(\boldsymbol{\theta}) \tag{2.38}$$

2.1.3 E and B mode decomposition

The shear tensor Γ can be expressed as a sum of two components corresponding to the number of degrees of freedom. By introducing two scalar fields $\psi_E(\boldsymbol{\theta})$ and $\psi_B(\boldsymbol{\theta})$, we decompose the shear matrix Γ_{ij} $(i, j = 1, 2)$ into two independent modes as (Crittenden et al., 2002):

$$\Gamma(\boldsymbol{\theta}) = \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{pmatrix} = \Gamma^{(E)}(\boldsymbol{\theta}) + \Gamma^{(B)}(\boldsymbol{\theta}), \tag{2.39}$$

with

$$\Gamma^E_{ij}(\boldsymbol{\theta}) = \left(\partial_i\partial_j - \delta_{ij}\frac{1}{2}\nabla^2\right)\psi_E(\boldsymbol{\theta}), \tag{2.40}$$

$$\Gamma^B_{ij}(\boldsymbol{\theta}) = \frac{1}{2}(\epsilon_{kj}\partial_i\partial_k + \epsilon_{ki}\partial_j\partial_k)\psi_B(\boldsymbol{\theta}), \tag{2.41}$$

where ϵ_{ij} is the Levi-Civita symbol in two dimensions, defined such that $\epsilon_{11} = \epsilon_{22} = 0$, $\epsilon_{12} = -\epsilon_{21} = 1$. Here the first term associated with ψ_E is a gradient or scalar E component and the second term with ψ_B is a curl or pseudoscalar B component. The shear components (γ_1, γ_2) are written in terms of ψ_E and ψ_B as:

$$\gamma_1 = +\Gamma_{11} = -\Gamma_{22} = \frac{1}{2}(\psi_{E,11} - \psi_{E,22}) - \psi_{B,12} \tag{2.42}$$

$$\gamma_2 = \Gamma_{12} = \Gamma_{21} = \psi_{E,12} + \frac{1}{2}(\psi_{B,11} - \psi_{B,22}). \tag{2.43}$$

Now comparing the expression of the E-mode of the shear tensor in Eq. 2.40 with the shear tensor in the general index notation in Eq. 2.23, and the two shear components γ_1 and γ_2 in Eq. 2.42 and Eq. 2.43 were with the E-B components of ψ with the corresponding general expressions in Eq. 2.18 and 2.19 we find that $\psi_E(\boldsymbol{\theta}) = \psi(\boldsymbol{\theta})$ and $\psi_B(\boldsymbol{\theta}) = 0$. Hence, for a lensing-induced shear field, the E-mode signal is related to the convergence κ, i.e. the surface mass density of the lens, while the B-mode signal is identically zero. Figure 2.3 illustrates characteristic distortion patterns from E-mode (curl-free) and B-mode (divergence-free) fields. Weak lensing only produces curl-free E-mode signals so that the presence of divergence-free B-modes

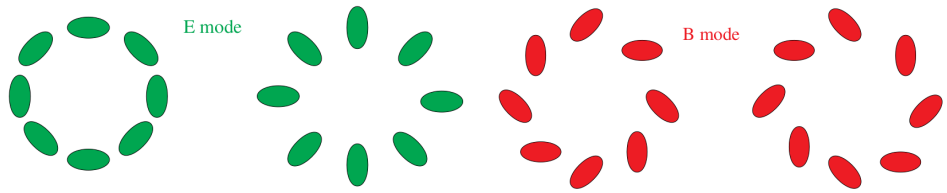


Figure 2.3 *Illustration of shape distortion patterns from E-mode and B-mode fields. The distortion pattern from E-mode is perfectly radial or tangential (Umetsu, 2020).*

can be used as a null test for systematic effects. In the weak-lensing regime, a tangential E-mode pattern is produced by a positive mass overdensity (e.g. halos), while a radial E-mode pattern is produced by a negative mass overdensity (e.g. cosmic voids).

Umetsu (2020) shows that gravitational lensing can induce B-modes, for example, when multiple deflections of light are involved (multiple lenses on the line of sight). However, these B-modes can be generated at higher orders, and the B-mode contributions coming from multiple deflections are suppressed by a large factor compared to the E-mode contributions. In real observations, intrinsic ellipticities of background galaxies also contribute to weak-lensing shear estimates. Assuming that intrinsic ellipticities have random orientations in projected space, such an isotropic ellipticity distribution will yield statistically identical contributions to E and B-modes. Therefore, the B-mode signal provides a useful null test for systematic effects in weak-lensing observations.

2.2 A weak lensing estimator: the aperture mass

From Eq. 2.31, we know that is possible to express shear as a convolution of convergence. This can be inverted with appropriate *inversion algorithms* (Bartelmann and Schneider (2001); Kaiser and Squires (1993); Schneider (1996)).
Considering the Fourier transform of the convergence κ and the shear convolution kernel $\mathcal{D}$ defined as Eq. 2.33:

$$\hat{\kappa}(\boldsymbol{l}) = \int_{\mathbb{R}^2} e^{i\boldsymbol{\theta}\cdot\boldsymbol{l}}\kappa(\boldsymbol{\theta})\mathrm{d}^2\boldsymbol{\theta}, \tag{2.44}$$

$$\hat{\mathcal{D}}(\boldsymbol{l}) = \pi\frac{l_1^2 - l_2^2 + 2il_1l_2}{|\boldsymbol{l}|^2}, \tag{2.45}$$

$$\hat{\mathcal{D}}^*(\boldsymbol{l}) = \pi\frac{l_1^2 - l_2^2 - 2il_1l_2}{|\boldsymbol{l}|^2}, \tag{2.46}$$

let us note that $\hat{\mathcal{D}}\mathcal{D}^* = \pi^2$, where the asterisk denotes again the complex conjugate. Using the properties of the Fourier transform regarding the convolution, we have that $\hat{\gamma}(\boldsymbol{l}) = \pi^{-1}\hat{\kappa}(\boldsymbol{l})\hat{\mathcal{D}}()$, from which we obtain $\hat{\kappa}(\boldsymbol{l}) = \frac{1}{\pi}\hat{\gamma}(\boldsymbol{l})\hat{\mathcal{D}}^*(\boldsymbol{l})$. In the real space this becomes:

$$\kappa(\boldsymbol{\theta}) - \kappa_0 = \frac{1}{\pi}\int_{\mathbb{R}^2}\gamma(\boldsymbol{\theta'})\mathcal{D}^*(\boldsymbol{\theta}-\boldsymbol{\theta'})\mathrm{d}^2\boldsymbol{\theta'} = \frac{1}{\pi}\int_{\mathbb{R}^2}\Re[\gamma(\boldsymbol{\theta'})\mathcal{D}^*(\boldsymbol{\theta}-\boldsymbol{\theta'})]\mathrm{d}^2\boldsymbol{\theta'}. \tag{2.47}$$

With the constant κ_0 we can show that the shear does not change if we add a sheet of constant convergence (i.e. *mass sheet degeneracy*). The symbol $\Re$ refers to the real part of the complex number, and the equality comes from the constraint that the convergence is a real function. In practice, it is possible to use this relation to estimate the convergence by using a dataset of measured complex ellipticities $\epsilon_i = \epsilon(\boldsymbol{\theta_i})$ instead of the shear (equivalent to the reduced shear in the weak-lensing limit). the estimator will be:

$$\kappa(\boldsymbol{\theta}) = \frac{1}{N\pi} \sum_{i=1}^{N} \Re[\epsilon_i \mathcal{D}^*(\boldsymbol{\theta} - \boldsymbol{\theta}_i)]. \tag{2.48}$$

This is one of the most basic estimators to reconstruct mass surface density from shear measurements and analyze it to detect mass concentrations, but the relations between galaxy ellipticities and the resulting mass maps are generally complicated.

The estimator we adopt in this Thesis and the most common one is the *Aperture Mass*

$$M_{ap}(\boldsymbol{\theta}) = \int_{\mathbb{R}^2} \kappa(\boldsymbol{\theta'}) U(\boldsymbol{\theta} - \boldsymbol{\theta'}) \mathrm{d}^2\boldsymbol{\theta'}. \tag{2.49}$$

Aperture mass is defined as the convolution of a convergence map with a weight function $U(\boldsymbol{\theta} - \boldsymbol{\theta'})$; the weight function can be selected as an "aperture function" that tends to zero when the distance $|\boldsymbol{\theta} - \boldsymbol{\theta'}|$ increases. We have then that the aperture mass becomes an average estimate of the convergence weighted within the aperture with the weight function U. If we assume that the aperture is symmetric $U(\boldsymbol{\theta} - \boldsymbol{\theta'}) = U(|\boldsymbol{\theta} - \boldsymbol{\theta'}|)$. By inserting Eq. 2.48 in Eq. 2.49 we get:

$$\begin{aligned} M_{ap}(\boldsymbol{\theta}) &= \int_{\mathbb{R}^2} U(|\boldsymbol{\theta} - \boldsymbol{\theta'}|) \left[\frac{1}{\pi} \int_{\mathbb{R}^2} \Re[\gamma(\boldsymbol{\theta''}) \mathcal{D}^*(\boldsymbol{\theta'} - \boldsymbol{\theta''})] \mathrm{d}^2\boldsymbol{\theta''} + \kappa_0 \right] \mathrm{d}^2\boldsymbol{\theta'} \\ &= \frac{1}{\pi} \Re \int_{\mathbb{R}^2} \gamma(\boldsymbol{\theta''}) \int_{\mathbb{R}^2} U(|\boldsymbol{\theta} - \boldsymbol{\theta'}|) \mathcal{D}^*(\boldsymbol{\theta'} - \boldsymbol{\theta''}) \mathrm{d}^2\boldsymbol{\theta'} \mathrm{d}^2\boldsymbol{\theta''} + \\ &+ \kappa_0 \int_{\Re^2} U(|\boldsymbol{\theta} - \boldsymbol{\theta'}|) \mathrm{d}^2\boldsymbol{\theta'}. \end{aligned} \tag{2.50}$$

The second integral can be written in polar coordinates as

$$\kappa_0 \int_{\mathbb{R}^2} U(|\boldsymbol{\theta} - \boldsymbol{\theta'}|) \mathrm{d}^2\boldsymbol{\theta'} = \kappa_0 2\pi \int_0^\infty U(x) x \mathrm{d}x. \tag{2.51}$$

Let us suppose that the weight function U is *compensated*, that is

$$\int_0^\infty U(x) x \mathrm{d}x = 0. \tag{2.52}$$

In this way the aperture mass will not depend on κ_0; it can be expressed as:

$$\begin{aligned} M_{ap}(\boldsymbol{\theta}) &= \frac{1}{\pi} \Re \int_{\mathbb{R}^2} \gamma(\boldsymbol{\theta''}) \int_{\mathbb{R}^2} U(|\boldsymbol{\theta} - \boldsymbol{\theta'}|) \mathcal{D}^*(\boldsymbol{\theta'} - \boldsymbol{\theta''}) \mathrm{d}^2\boldsymbol{\theta'} \mathrm{d}^2\boldsymbol{\theta''} = \\ &= \frac{1}{\pi} \Re \int_{\mathbb{R}^2} \gamma(\boldsymbol{\theta''}) \int_{\mathbb{R}^2} U(|\boldsymbol{\theta'}|) \mathcal{D}^*(\boldsymbol{\theta'} - \boldsymbol{\theta''} + \boldsymbol{\theta}) \mathrm{d}^2\boldsymbol{\theta'} \mathrm{d}^2\boldsymbol{\theta''} = \\ &= \frac{1}{\pi} \Re \int_{\mathbb{R}^2} \gamma(\boldsymbol{\theta''}) \int_0^\infty U(x) x \int_0^{2\pi} \mathcal{D}^*(X - Y) \mathrm{d}\phi \mathrm{d}x \mathrm{d}^2\boldsymbol{\theta''} = \\ &= -\frac{1}{\pi} \Re \int_{\mathbb{R}^2} \gamma(\boldsymbol{\theta''}) \int_0^\infty U(x) x \int_0^{2\pi} \frac{1}{(xe^{i\phi} - Y)^2} \mathrm{d}\phi \mathrm{d}x \mathrm{d}^2\boldsymbol{\theta''}, \end{aligned} \tag{2.53}$$

with the conplex numbers $X = \theta_1' + i\theta_2' = xe^{i\phi}$ and $Y = (\theta_1'' - \theta_1) + i(\theta_2'' - \theta_2)$. The complex integral is solved by using $\mathrm{d}\phi = -i\frac{\mathrm{d}X}{X}$, transforming it into a loop integral and using the residual theorem:

$$\int_0^{2\pi} \frac{1}{(xe^{i\phi} - Y)^2}\mathrm{d}\phi = \frac{\pi}{Y^2}[2\Theta(|Y| - |X|) - |Y|\delta(|X| - |Y|)], \tag{2.54}$$

where Θ is Heaviside's step function and δ is Dirac's delta. The aperture mass becomes:

$$\begin{aligned}
M_{ap}(\boldsymbol{\theta}) &= \frac{1}{\pi}\Re\int_{\mathbb{R}^2}\gamma(\boldsymbol{\theta''})\int_0^\infty \frac{U(x)x\pi}{Y^2}[2\Theta(|Y| - |X|) - |Y|\delta(|X| - |Y|)]\mathrm{d}x\mathrm{d}^2\boldsymbol{\theta''} = \\
&= \Re\int_{\mathbb{R}^2}\frac{\gamma(\boldsymbol{\theta''})}{Y^2}\left[|Y|^2U(|Y|) - 2\int_0^{|Y|}U(x)x\mathrm{d}x\right]\mathrm{d}^2\boldsymbol{\theta''} = \\
&= \Re\int_{\mathbb{R}^2}\frac{\gamma(\boldsymbol{\theta}+\boldsymbol{\theta'})}{(\theta')^2}\left[|\boldsymbol{\theta'}|^2U(|\boldsymbol{\theta'}|) - 2\int_0^{|\boldsymbol{\theta'}|}U(x)x\mathrm{d}x\right]\mathrm{d}^2\boldsymbol{\theta'} = \\
&= \int_{\mathbb{R}^2}\left[-|\boldsymbol{\theta'}|^2\frac{\gamma(\boldsymbol{\theta}+\boldsymbol{\theta'})}{(\theta')^2}\right]\left[\frac{2}{|\boldsymbol{\theta'}|^2}\int_0^{|\boldsymbol{\theta'}|}U(x)x\mathrm{d}x - U(|\boldsymbol{\theta'}|)\right]\mathrm{d}^2\boldsymbol{\theta'} = \\
&= \int_{\mathbb{R}^2}\left[-|\boldsymbol{\theta'}-\boldsymbol{\theta}|^2\Re\left(\frac{\gamma(\boldsymbol{\theta'})}{(\theta'-\theta)^2}\right)\right]\left[\int_0^{|\boldsymbol{\theta'}-\boldsymbol{\theta}|}\frac{2U(x)x}{|\boldsymbol{\theta'}-\boldsymbol{\theta}|^2}\mathrm{d}x - U(|\boldsymbol{\theta'}-\boldsymbol{\theta}|)\right]\mathrm{d}^2\boldsymbol{\theta'},
\end{aligned} \tag{2.55}$$

where we have introduced the complex numbers $\theta' = \theta_1' + i\theta_2'$ and $\theta = \theta_1 + i\theta_2$ corresponding to the vectors $\boldsymbol{\theta'}$ and $\boldsymbol{\theta}$. The aperture mass can be written in a more immediate way if we define the *tangential shear component at the position* $\boldsymbol{\theta'}$ *with respect to* $\boldsymbol{\theta}$:

$$\gamma_+(\boldsymbol{\theta'};\boldsymbol{\theta}) = -|\boldsymbol{\theta'}-\boldsymbol{\theta}|^2\Re\left(\frac{\gamma(\boldsymbol{\theta'})}{(\theta'-\theta)^2}\right) = -\Re\left(\frac{\gamma(\boldsymbol{\theta'})(\theta'-\theta)^*}{(\theta'-\theta)}\right), \tag{2.56}$$

and a *filter function*, a real function of a real variable t:

$$Q(t) = \frac{2}{t^2}\int_0^t U(t')t'\mathrm{d}t' - U(t). \tag{2.57}$$

The aperture mass can thus be written as:

$$M_{ap}(\boldsymbol{\theta}) = \int_{\mathbb{R}^2}\gamma_+(\boldsymbol{\theta'};\boldsymbol{\theta})Q(|\boldsymbol{\theta'}-\boldsymbol{\theta}|)\mathrm{d}^2\boldsymbol{\theta'}. \tag{2.58}$$

This equation allows us to express the aperture mass, a $U-$weighted integral of the convergence, as an integral over the tangential shear weighted by the filter function Q. Different filters have different properties; generally, they have compact support so that $M_{ap}(\boldsymbol{\theta})$ can be expressed as an integral of the tangential shear over a finite area around $\boldsymbol{\theta}$, and with a suppression in the aperture center so that the tangential shear does not diverge and the weak lensing approximation ($\kappa \ll 1, |\gamma| \ll 1$) does not break down. It is generally possible to create a signal-to-noise ratio map of the mass aperture statistic with a chosen filter Q, in which sufficiently strong lenses will appear as significant peaks. This method has been tested on synthetically generated

data, and it is shown in several works that it lives up to the expectations (Maturi et al. (2005); Pace et al. (2007)). The signal-to-noise ratio map is not corrupted by small-scale deflectors (such as individual galaxies), nor by large-scale deflectors. This method can then be used to search for mass concentrations (dark halos of galaxy clusters) given wide-field images of ellipticities.

2.2.1 Tangential and cross shear components

As we have seen in Section 2.1.2, the spin-2 shear components γ_1 and γ_2 are coordinate-dependent, defined relative to a reference Cartesian coordinate frame. It is useful to consider components of the shear that are coordinate-independent with respect to a given reference point, such as the cluster center. We already introduced in Eq. 2.56 the *tangential shear component* as the real part of a given complex number. If we consider the imaginary part, we can define the equivalent *cross shear component at position* $\boldsymbol{\theta'}$ *with respect to* $\boldsymbol{\theta}$:

$$\gamma_\times(\boldsymbol{\theta'};\boldsymbol{\theta}) = -|\boldsymbol{\theta'}-\boldsymbol{\theta}|^2\Im\left(\frac{\gamma(\boldsymbol{\theta'})}{(\theta'-\theta)^2}\right) = -\Im\left(\frac{\gamma(\boldsymbol{\theta'})(\theta'-\theta)^*}{(\theta'-\theta)}\right). \tag{2.59}$$

The tangential and cross shear components can also be expressed as explicit functions of the cartesian components of $\Delta\theta = \theta' - \theta = \Delta\theta_1 + i\Delta\theta_2$ as:

$$\gamma_+(\boldsymbol{\theta'};\Delta\theta) = -\frac{1}{\Delta\theta_1^2+\Delta\theta_2^2}\left[\gamma_1(\boldsymbol{\theta'})(\Delta\theta_1^2-\Delta\theta_2^2) + 2\gamma_2(\boldsymbol{\theta'})(\Delta\theta_1\Delta\theta_2)\right] \tag{2.60}$$

$$\gamma_\times(\boldsymbol{\theta'};\Delta\theta) = -\frac{1}{\Delta\theta_1^2+\Delta\theta_2^2}\left[\gamma_2(\boldsymbol{\theta'})(\Delta\theta_1^2-\Delta\theta_2^2) - 2\gamma_1(\boldsymbol{\theta'})(\Delta\theta_1\Delta\theta_2)\right] \tag{2.61}$$

It is important to note that the cross shear is equivalent to computing the tangential shear on a map where each galaxy shear has been rotated by $\pi/4$, which according to Eq. 2.25 means using a new shear γ' whose components are:

$$\gamma_1' = \gamma_2, \qquad \gamma_2' = -\gamma_1. \tag{2.62}$$

The tangential and cross shear can also be expressed as a function of the radial coordinates of $\Delta\theta = \vartheta(\cos\phi + i\sin\phi)$ as:

$$\begin{cases} \gamma_+'(\boldsymbol{\theta'};\phi) = +\gamma_1(\boldsymbol{\theta'})\cos 2\phi + \gamma_2(\boldsymbol{\theta'})\sin 2\phi \\ \gamma_\times'(\boldsymbol{\theta'};\phi) = -\gamma_1(\boldsymbol{\theta'})\sin 2\phi + \gamma_2(\boldsymbol{\theta'})\cos 2\phi \end{cases}. \tag{2.63}$$

These quantities are directly observable in the weak-lensing limit where $\kappa \ll 1$, $\gamma \ll 1$. Let us fix the reference point of the tangential and cross components $\boldsymbol{\theta}$ and let us compute their average value on a circle around $\boldsymbol{\theta'}$ that has a fixed radius ϑ. Using the two-dimensional version of Gauss' theorem, it is possible to show:

$$\langle\gamma_+\rangle(\vartheta;\boldsymbol{\theta}) = \frac{1}{2\pi}\oint \gamma_+(\theta+\vartheta e^{i\phi})\mathrm{d}\phi = \frac{\bar{\Sigma}-\Sigma}{\Sigma_{crit}} = \frac{\Delta\Sigma}{\Sigma_{crit}}(\vartheta;\boldsymbol{\theta}) \tag{2.64}$$

$$\langle\gamma_\times\rangle(\vartheta;\boldsymbol{\theta}) = \frac{1}{2\pi}\oint \gamma_+(\theta+\vartheta e^{i\phi})\mathrm{d}\phi = 0, \tag{2.65}$$

where $\Delta\Sigma(\vartheta;\boldsymbol{\theta})$ is the excess surface mass density around $\boldsymbol{\theta}$; it is the azimuthally averaged surface density profile in a circle of radius ϑ around $\boldsymbol{\theta}$ (Umetsu, 2020).

Eq. 2.65 shows that the cross shear component averaged around the loop measures the B-mode distortion pattern since it is expected to be statistically consistent with zero if the signal is due to weak lensing. Therefore, a measurement of the B-mode signal provides a useful null test against systematic errors. Eq. 2.64 shows that given an arbitrary circular loop of radius around the chosen center $\boldsymbol{\theta}$, the tangential shear component averaged around the loop extracts a signal that is only due to the convergence; it thus extracts the E-mode distortion pattern (see Section 2.1.3). It is interesting to notice that if one rotates its shear map by $\pi/4$ the E-mode and B-mode are swapped.

2.2.2 Observable galaxy ellipticities

In the real Universe the background sources of gravitational lensing events are never circular, thus the distortion observed is partially due to weak lensing and partially due to their intrinsic shape. In the case of galaxies, we can define an intrinsic complex ellipticity $\epsilon^{(s)}$.
There is a link between the reduced shear g, the intrinsic complex ellipticity $\epsilon^{(s)}$ for a source and the complex ellipticity of its image ϵ. For an elliptical source with axis ratio $r \leq 1$ the complex ellipticity has modulus $|\epsilon^{(s)}| = \frac{|1-r|}{|1+r|}$ and a phase given by twice the angle between the major axis and the positive θ_1-direction. Schneider and Seitz (1994) showed that the relationship between intrinsic and observed complex ellipticities is

$$\epsilon^{(s)} = \begin{cases} \dfrac{\epsilon - g}{1 - g^*\epsilon} & \text{if} \quad |g| \leq 1, \\ \dfrac{1 - g\epsilon^*}{\epsilon^* - g^*} & \text{if} \quad |g| > 1. \end{cases} \tag{2.66}$$

The inverse transformation is obtained by interchanging the source and the image ellipticity and making the substitution $g \to -g$. If we make the assumption that the source galaxies are randomly oriented, so that $\langle \epsilon^{(s)} \rangle = 0$, when averaging over a sufficiently large number of them, then Eq. 2.66 tells us that

$$\langle \epsilon \rangle = \begin{cases} g & \text{if} \quad |g| \leq 1, \\ 1/g^* & \text{if} \quad |g| > 1, \end{cases} \tag{2.67}$$

i.e. the expectation value of the image ellipticity is the reduced shear. If we consider the sub-critical lenses in the weak lensing approximation $\kappa \ll 1$:

$$\langle \epsilon \rangle = g \approx \gamma. \tag{2.68}$$

Thus, in the case of weak lensing, the ellipticity of a galaxy image is an unbiased estimate of the local shear plus the intrinsic shape dispersion of background galaxies that we can assume null almost always.

Let us now suppose we have a dataset of complex galaxy ellipticities ϵ_i with positions $\boldsymbol{\theta_i}$, with a number density of galaxy images n_g. We can define the observational version of Eq. 2.56 as the *complex tangential ellipticity of the image in position* $\boldsymbol{\theta_i}$ *with respect to* $\boldsymbol{\theta}$ as

$$\epsilon_+(\boldsymbol{\theta_i}; \boldsymbol{\theta}) = \epsilon_{+i}(\boldsymbol{\theta}) = -\Re\left(\frac{\epsilon_i(\theta_i - \theta)^*}{(\theta_i - \theta)}\right) , \tag{2.69}$$

and the observational estimator of the mass aperture 2.58 as:

$$M_{ap}(\boldsymbol{\theta}) = \frac{1}{n_g} \sum_i \epsilon_{+i}(\boldsymbol{\theta}) Q(|\boldsymbol{\theta_i} - \boldsymbol{\theta}|), \tag{2.70}$$

where the sum extends over all galaxy images within the aperture, i.e., within the support of the filter Q (where $Q \neq 0$). It can be shown that the variance of this estimator is given by:

$$\sigma^2_{M_{ap}}(\boldsymbol{\theta}) = \frac{1}{2n_g^2} \sum_i |\epsilon_{+i}(\boldsymbol{\theta})|^2 Q^2(|\boldsymbol{\theta_i} - \boldsymbol{\theta}|). \tag{2.71}$$

The noise produced by the intrinsic galaxy ellipticities can be taken as the square root of this variance, and the signal-to-noise ratio can then be computed; it is independent of the normalization of the weight function.

Chapter 3

Galaxy clusters: detection and their use in comsology

The *hierarchical model* of cosmic structure formation is at the moment considered the standard model. As described in Section 1.2 the small overdensities are the first to collapse against the cosmic expansion and form the first "halos". Then, through a sequence of mergers and aggregation, driven only by gravity and dark matter, bigger structures are formed. Galaxy clusters are the results of this process given by the aggregation of the smaller galaxy-size halos. They are the biggest virialized structures in the Universe and thus they represent the peaks of the dark matter density field.

These structures can include different amounts of galaxies spanning from a few to thousands. In literature, aggregations of galaxies are commonly called *clusters* or *groups*. A clear threshold does not univocally define the distinction between the two, but it usually consists of a difference in mass or richness, i.e. the number of galaxies in the structure. Assemblies of galaxies that have characteristic masses lower than $10^{14}M_{\odot}$ and host less than ~ 50 member galaxies have characteristic size $\lesssim 1$ Mpc and are usually referred to as *groups*.

Zwicky noted in 1933 that the galaxies in the Coma cluster and other rich clusters move so fast that the clusters required to keep the galaxies bound about 10 to 100 times more mass than the luminous galaxies themselves could account for that. This was the earliest indication that there is invisible mass, or dark matter, in at least some objects in the Universe.

When X-ray telescopes became available after 1966, it was discovered that clusters emit $\approx 10^{13-15}\text{ergs}^{-1}$. The source of this powerful X-ray emission (thermal bremsstrahlung, free-free radiation) is a hot, plasma with temperature in the range 10^{7-8}K and density of 10^{-3}cm^{-3}. Based on the assumption that this intra-cluster gas is in hydrostatic equilibrium with a spherically symmetric gravitational potential provided by the total cluster matter, the X-ray temperature and flux can be used to estimate the cluster mass. Typical results approximately agree with the mass estimates from the kinematics of cluster galaxies exploiting the virial theorem. Today we know that about 85% of the cluster's mass is dark matter, and about 15% is the intra-cluster medium ICM (the hot gas). The X-ray emission thus independently confirms the existence of dark matter in galaxy clusters.

Later, luminous arc-like features were discovered in galaxy clusters. Their light

is typically bluer than that from the cluster galaxies, and their length is comparable to the size of the central cluster region. These arcs are the signatures of *strong gravitational lensing*, which we briefly introduced in Chapter 2. They are the image of background galaxies that are strongly distorted by the gravitational tidal field close to the centres of the clusters. At larger distances from the centre, the clusters can only weakly deform images of background galaxies since the lens is subcritical and the effect enters the *weak regime*.

In Section 3.1 we will explain some aspects of the exploitation of clusters in cosmology and why cluster studies through weak lensing measures are crucial (Bartelmann and Schneider (2001); Schneider (2006)). Then Section 3.2 will present an overview of the methods of clusters' detection with a focus on weak-lensing detection in Section 3.2.4 with main reference Bartelmann and Maturi (2016).

3.1 Cluster cosmology

In Chapter 1, we described the evolution of the density fluctuations through the history of the Universe. However, we focused only on the concepts that come up from the linear approach of the perturbation theory, which is valid while $\delta < 1$, which is only true for the early stages of cosmic history or very large scales today. At late stages of cosmic history and on small scales, the growth of density fluctuations begins to depart from the linear regime. When density fluctuations grow in a non-linear way, fluctuations of different sizes interact, and the evolution of $\mathcal{P}(k)$ becomes so complicated that it needs to be evaluated numerically. The non-linear growth of the density fluctuations is crucial for accurately calculating weak lensing effects by large-scale structures and, thus, crucial to this work. When $\delta > 1$, the perturbation becomes non-linear. The fluctuations are no longer fully characterized by the power spectrum, as seen in Section 1.2.2 only because they develop non-Gaussian features. The non-linear regime is approached first by the small-scale perturbations (large k). In Eq. 1.40 we showed that $\sigma^2 \sim \delta^2$, so when $\delta \sim 1$ then also $\sigma^2 \sim 1$. We can obtain from the Eq. 1.40 a characteristic scale length, wave number or mass, the *non-linear scale*, at which the variance of matter density contrast becomes 1, at each cosmological epoch. The mass scale is:

$$M \sim \delta_+^{\frac{6}{n+3}}(t). \tag{3.1}$$

Non-linear perturbations evolved to form the structures we observe in the present-day Universe, the largest of which are the clusters of galaxies. They hold then a peculiar role in large-scale structures; thus they are very powerful tools for cosmological studies; they depend only weakly on baryon physical processes but they are closely coupled with extreme sensitivity to the expansion history of the Universe and structure formation as they are the ultimate product of the hierarchical standard model (Allen, Evrard, and Mantz, 2011).

We already said that the normalization of the power spectrum is not fixed by theoretical models but must be measured from observations. One way is to measure the normalization of the anisotropies of the CMB. Another way involves the (mass) variance of the density fluctuations. With this method, the normalization is described by the well-known cosmological parameter σ_8. The mass variance is not directly observable, but we can pass through the variance of galaxy counts: in fact, most of the matter, i.e. the dark matter, is not visible, and therefore the spatial

clustering of the observable matter is not directly described by the matter power spectrum $\mathcal{P}_m(k,a)$ but by a modified or *biased* version of it (Kaiser, 1984):

$$\mathcal{P}_{obs}(k,a) = b^2(M,a)\mathcal{P}_m(k,a), \tag{3.2}$$

where the *bias factor* $b(M,a)$ accounts for this modification that depends on mass and cosmic epoch. By measuring the local variance of galaxy counts within certain volumes and assuming an expression for $b(M,a)$, the amplitude of dark matter fluctuations can be inferred. Conventionally it is measured $\sigma_{8,gal}$ whithin spheres of $8h^{-1}$Mpc, and the result is $\sigma_{8,gal} \approx 1$ at the present time. Then with the expression of the bias it is possible to pass from $\sigma_{8,gal}$ to the variance σ_8 of matter, thus discovering the power spectrum normalization.

More than galaxies, clusters of galaxies are highly biased, and the power spectrum associated with them results strongly enhanced over the matter spectrum (Allen et al. (2011);Desjacques et al. (2018); Tinker et al. (2010)).

The statistical study of the distribution of structures relies on the mathematical background given by the Press-Schecter formalism Press and Schechter (1974), which has as a core concept the *mass function*. This function, expressing the mean number density of a population of structures with a given mass per co-moving volume, is given by

$$\frac{\mathrm{d}n}{\mathrm{d}\ln M} = \frac{\bar{\rho}}{M} f(\sigma) \left| \frac{\mathrm{d}\ln\sigma}{\mathrm{d}\ln M} \right|, \tag{3.3}$$

where $\bar{\rho}$ is the mean background matter density, σ^2 the variance and $f(\sigma)$ a model-dependent function of it, also known as *multiplicity function*.

It is possible then to connect theory with observations by introducing actual counts of clusters, the main goal of most of the modern wide clusters surveys like Euclid or KiDS. The counts can be expressed as N, the number of clusters expected for a bin in mass and redshift in a solid angle Ω:

$$N(M,z) = \frac{\Omega}{4\pi} \int \mathrm{d}z \frac{\mathrm{d}V}{\mathrm{d}z} \int \mathrm{d}\ln M \frac{\mathrm{d}n}{\mathrm{d}\ln M}, \tag{3.4}$$

where the cosmological background comes into play both in the $\frac{\mathrm{d}V}{\mathrm{d}z}$ term, the co-moving volume element that has a dependence on the scale factor, and of course with the mass function 3.3 $\frac{\mathrm{d}n}{\mathrm{d}\ln M}$. The mass function is sensitive to the density parameter of matter Ω_m and on σ that constrains the parameters of the linear power spectrum.

Galaxy clusters, as the most massive bound structures in the LSS, define the high-mass end of the mass function and are, therefore, extremely sensitive to the choice of the cosmological model. In complementarity with CMB observations, galaxy clustering and weak lensing measurements, studying the abundance of galaxy clusters provides a unique tool to constrain cosmological parameters. The best-constrained parameters from cluster studies, Ω_m and σ_8 (e.g. Abdullah, Klypin, and Wilson (2020); Lesci et al. (2022); Sartoris et al. (2016)) have degeneracies that are different with respect to other methods, and therefore they help to interpret parameter measurements and uncertainties (e.g. Rosati, Borgani, and Norman (2002); Rozo et al. (2009); Vikhlinin et al. (2009)).

3.2 Cluster detection

There are several methods to detect galaxy clusters in observational data that depend on the different properties and components (e.g. ICM, dark matter, galaxies, etc.) of these peculiar objects. Here we describe the main methods to accomplish this task.

3.2.1 Optical and NIR observations

Galaxy clusters can be detected from the observation of the galaxies they are composed of through optical and near-infrared (NIR) wavelengths. This technique is subject to projection effects, for which field galaxies along the line of sight could be mistakenly identified as members of a cluster.

A solution to this problem is to exploit the photometric properties characterizing the galaxies which are members of a cluster. Galaxy clusters contain a well-defined, highly regular population of elliptical and lenticular galaxies, observed in optical and NIR bands. These member galaxies in a color-magnitude diagram dispose along a sequence, called *red-sequence* (Bower et al., 1992). In particular, there is a tight linear relationship between color and magnitude, showing that the redder galaxies are brighter. Therefore this relation plays a considerable role in the identification of galaxy clusters.

There are other detection algorithms that for optical and NIR observations that do not rely on the red-sequence. For example, the AMICO algorithm (Bellagamba et al. (2011); Bellagamba et al. (2018)) assumes a model for the density and luminosity profiles of galaxy clusters and identifies them by finding galaxy overdensities and matching them with the defined model. In particular, in this case, each galaxy has an assigned probability to be a member of a cluster or to be part of the field. In this Thesis, we use the structure of this technique and adapt it for weak lensing data analysis.

3.2.2 X-ray observations

Despite galaxy clusters are optically localized via the emission of the galactic component, in terms of mass the dominant baryonic component is the ICM. Given its high temperature, the ICM is responsible for a powerful thermal free-free bremsstrahlung emission. This is the main origin of the typical luminosities of the order of 10^{43} erg s$^{-1} \leq L_X \leq 10^{45}$ erg s^{-1} (Rosati, Borgani, and Norman, 2002).

Observationally speaking, the X-ray emission of galaxy clusters is a powerful tool to detect them. Moreover, X-ray observations also give the opportunity to carry out spectral analyses that shed light on the density and temperature distributions of clysters and their metal content (thanks to the presence of emission lines in the X-ray spectra).

The main advantages of the X-ray identification of clusters are the availability of sufficiently precise and simple relations between observables (e.g. luminosity, temperature) and mass (Mantz et al., 2010) and the fact that is a technique nearly insensitive to projection effect, for the nature of the emission itself (Allen, Evrard, and Mantz, 2011).

3.2.3 Sunyaev Zel'dovich effect

Galaxy clusters are also detectable in the microwaves (Bleem et al., 2015), thanks to the so-called Sunyaev and Zeldovich effect (1970), SZ for short. The effect consists basically of the Inverse Compton between the CMB photons and the free electrons in the ICM. The CMB radiation is well described by a black-body emission with temperature $T_{CMB} \simeq 2.726$ K (Fixsen, 2009). The CMB photons are much less energetic than the cluster electrons, so the interaction between the two and the consequent energy gain performed by the CMB photon become visible as a distortion in the black-body CMB spectrum at the respective cluster location. In particular, this distortion consists of a shift with a consequent lower intensity for frequencies $\nu < 217$ GHz and higher intensity for $\nu > 217$ GHz.

The big advantage of the SZ detection method is that is very weakly dependent on redshift, being the source of photons in the CMB, so it is particularly suited for the detection of massive high-z clusters but it is less sensitive to low masses than X-ray and optical detections. The main observable is the Compton parameter y, which is proportional to gas density and temperature.

3.2.4 Gravitational lensing

Galaxy clusters, as the largest self-gravitating systems in the universe, act as powerful cosmic lenses, producing a variety of detectable lensing effects from strong to weak lensing, including deflection, shearing, and magnification of the images of background sources as described in Chapter 2. The critical advantage of cluster gravitational lensing is its ability to study the mass distribution of individual and ensemble systems independent of assumptions about their physical and dynamical state, as, for instance, it is for X-ray measurements (Bartelmann, 2010).

Galaxy clusters imprint a coherent weak distortion pattern onto the many faint and distant galaxies in their background. Since those distant galaxies reach number densities of $\simeq 40$ per square arc minute in typical images taken with large ground-based telescopes, typical galaxy clusters thus cover of order 10^3 background galaxies.

As shown in Chapter 2, shear and convergence are both related through the scalar lensing potential. Knowing the shear thus allows the scaled surface-mass density to be reconstructed. Cluster convergence maps can be obtained by convolving the measured shear with a simple kernel, opening the way to systematic, parameter-free, two-dimensional cluster studies (Kaiser and Squires (1993); Hoekstra (2014); Okabe and Umetsu (2008)). The weakness in this convolution algorithm, mainly due to the non-locality of the convolution, was identified and could be removed. The inversion techniques for matter distribution in galaxy-cluster lenses have been applied to numerous objects. For most of them, the mass-to-light ratios turned out to be $M/L \simeq (250 - 300)h\mathrm{M}_\odot/\mathrm{L}_\odot$ in blue and $M/L \simeq (150 - 200)h\mathrm{M}_\odot/\mathrm{L}_\odot$ in red light, respectively. Similar to weak lensing by galaxies, the masses quoted here are typically derived from parametrized density profiles adapted to the measured lensing signal and integrated to a fixed radius of order 1 h^{-1}Mpc. Statistically combining the weak-lensing signal of galaxy groups, the mass range in which mass-to-light ratios can be probed could be extended towards lower masses. For galaxy groups with masses around $\approx 10^{13-14}\mathrm{M}_\odot$, values of $M/L \simeq 180h\mathrm{M}_\odot/\mathrm{L}_\odot$ in blue and $M/L \simeq 250h\mathrm{M}_\odot/\mathrm{L}_\odot$ in red spectral ranges have been obtained.

Using a similar statistical analysis of gravitational lensing together with the

optical light distribution, the mass-to-light ratio of the central brightest galaxies in galaxy clusters was found to be $M/L \simeq 360h\mathrm{M}_\odot/\mathrm{L}_\odot$. Mass and light generally appear well correlated in weak lensing clusters.

Several detections of clusters with very high mass-to-light ratios have been claimed and raised the question of whether cluster-sized dark matter halos may exist which are so inefficient in producing stellar or X-ray emission that they are invisible to anything but gravitational lensing (e.g. Fischer (1999); Gray et al. (2001)).

Weak gravitational lensing also provides a powerful way to detect galaxy clusters regardless of their directly observable signatures. The *Aperture Mass* estimator, introduced in Section 2.2, is used in the methods developed for this purpose. Numerical simulations show that these methods are highly efficient in finding suitably massive matter concentrations if parameters and weight functions are optimally chosen to balance the completeness against the frequency of spurious detections carefully (Pace et al., 2007). One of the major challenges of this kind of technique is the requirement of separating the weak-lensing signal of dark-matter halos from that of the *large-scale structures* (LSS) they are embedded in. That is the goal of this Thesis, and the main feature of the algorithm we implement.

Chapter 4

The implementation of an optimal filter to weak lensing: AMICO-WL

The main purpose of this Thesis is to build an optimized method for weak-lensing detection of clusters of galaxies by extending the computational environment of AMICO (Adaptive Matched Identifier of Clustered Objects) (Bellagamba et al. 2011 (Bellagamba et al., 2011); 2018 (Bellagamba et al., 2018)). The code extension is called **AMICO-WL**, where WL stands for *weak-lensing*.

The core of AMICO is based on an Optimal Matched Filtering approach specialized for weak-lensing aperture mass measures (Bellagamba et al. (2011), 2018; Maturi et al. (2005)). An important advantage of the Optimal Matched Filtering formalism over other kinds of methods is its extreme flexibility, instead of being tailored for a specific signal and/or data sets (e.g. photometry, weak-gravitational lensing, SZ effect, and X-ray observations) for a wide range of redshifts. The algorithms developed should then weight the features in the data which are relevant for the cluster detection. The original implementation of AMICO can extract a specific signal from a set of data affected by noisy background, such as galaxy clusters or groups in a photometric catalog of galaxies, allowing to maximize the signal-to-noise ratio of the detected structures.
Another important feature of AMICO augmenting standard optimal filtering approaches is its ability to extract and remove the imprint of a detected cluster iteratively to highlight the presence of clusters with smaller *SNR* (signal-to-noise ratio) blended with other larger structures. This deblending technique is, of course, a point of crucial importance, especially in deep survey analysis, where a high number of detected clusters can be found per square degree.

AMICO has been successfully applied to photometric catalogs of galaxies for cosmological analyses (AMICO galaxy clusters in KiDS-DR3: Maturi et al. Maturi et al. (2019), Bellagamba et al. (2019); Giocoli et al. (2021); Cosmos: Toni et al. in prep.; HSC: Sartoris et al. in prep.) and simulations. As already mentioned, the versatility of Optimal Matched Filtering offers the possibility to implement other observables such as weak-gravitational lensing. In this work, we modified AMICO to adapt it to the gravitational lensing signal of galaxy clusters from the ellipticity

measurements of background galaxies. The filter was first developed by Maturi et al. (2005) where it was applied to simulated data and then applied to real data in Maturi et al. (2007).
Here, the filter assumes a mean radial profile of the halo shear pattern and a power spectrum for the noise; it is defined in the Fourier domain as proportional to their ratio so that the modes attributed to the clusters signal are amplified, and the modes due to the noise are suppressed. To separate dark matter halos from spurious peaks caused by the large-scale structure, the noise is composed of three contributions: Poisson noise from galaxy counts, noise from the intrinsic ellipticities of galaxies, and noise related to the weak lensing signal of the large scale structures.

The details of the optimal filter theory and the assumptions about the noise and the model adopted in the optimal filter will be discussed in Section 4.1 while Section 4.2 will present the full implementation of AMICO-WL.

4.1 The theory of the optimal filter

The matter distribution of the Universe, often described as the cosmic web, is composed of a continuous distribution of matter, the *large-scale structure* (LSS), usually with filamentary features, where the dark matter halos represent the peaks of mass density. In these peaks galaxy clusters are formed. When we measure the inhomogeneities of the projected mass distribution through weak lensing observations, i.e. galaxy ellipticity measures (see Section 2.2.2), the signal of halos is overlaid by the lensing signal of the LSS in front and behind the halos. In weak-lensing detections of clusters of galaxies, it is possible to mistakenly identify as clusters those peaks that are instead given by the large-scale structure lensing effect producing a spurious detection. It is not possible to strictly separate these two types of signal, however, a typical scale can be defined to discriminate theme. The large-scale structure can be considered as composed of dark-matter halos of broad and continuous mass range. Chapter 1 described the so-called *non-linear scale* as the scale at which the variance of the dark matter density contrast becomes unity: this scale separates the collapsed objects from the linear regime where the matter density can be considered as a linear superposition of linearly evolved perturbation modes.
Here, we model the contribution of virialized objects by assuming a radial profile for the shear signal and a power spectrum to represent the properties of the LSS. For the clusters, we assume the shear signal produced by a Navarro-Frenk-White (NFW) profile. For the large-scale structures, we adopt the expected shear power spectrum for a fiducial ΛCDM model. This definition was first used in Maturi et al. (2005) and Maturi et al. (2007) to construct a linear matched filter to detect the weak lensing signal of dark matter halos corresponding to galaxy clusters by suppressing the contamination of the large-scale structures, thus reducing the number of spurious detections.

The weak gravitational lensing signal of dark matter halo, $S(\boldsymbol{\theta})$, is given by an amplitude, A, and angular shape, $\tau(\boldsymbol{\theta})$. The data we measure, $D(\boldsymbol{\theta})$, are contaminated by the noise, $N(\boldsymbol{\theta})$, so that they can be written as:

$$D(\boldsymbol{\theta}) = S(\boldsymbol{\theta}) + N(\boldsymbol{\theta}) = A\tau(\boldsymbol{\theta}) + N(\boldsymbol{\theta}). \tag{4.1}$$

In this application, the signal will be the shear tangential components. The noise is made up of several contributions, which will be described in Section 4.1.2.

4.1.1 Definition of the optimal filter

The linear filter is defined as the convolution kernel, $\Psi(\boldsymbol{\theta})$, applied to the data, $D(\boldsymbol{\theta})$, such to yield an optimal estimate, A_{est}, of the amplitude, A, of the signal at the position $\boldsymbol{\theta}$,

$$A_{est}(\boldsymbol{\theta}) = \int_{\mathbb{R}^2} D(\boldsymbol{\theta'})\Psi(\boldsymbol{\theta} - \boldsymbol{\theta'})\mathrm{d}^2\boldsymbol{\theta'}. \tag{4.2}$$

This filter used in this linear estimator must satisfy two constraints. First, A_{est} must be unbiased, i.e its average over many realizations, b, has to vanish ($b = 0$):

$$b \equiv \langle A_{est} - A\rangle = A\left[\int_{\mathbb{R}^2} \tau(\boldsymbol{\theta'})\Psi(\boldsymbol{\theta'})\mathrm{d}^2\boldsymbol{\theta'}\right]. \tag{4.3}$$

Second, the measurement noise σ, determined by the mean-squared deviation of the estimate from its true value,

$$\begin{aligned}\sigma^2 = \langle(A_{est} - A)^2\rangle &= b^2 + \frac{1}{(2\pi)^2}\int_{\mathbb{R}^2} |\hat{\Psi}(\boldsymbol{k})|^2 P_N(k)\mathrm{d}^2\boldsymbol{k} \\ &= b^2 + \frac{1}{2\pi}\int_0^\infty |\hat{\Psi}(\mathbf{k})|^2 P_N(k)\mathrm{d}k\end{aligned} \tag{4.4}$$

has to be minimal. In order to find the function Ψ that satisfies these two conditions simultaneously, we combine them by means of a Lagrangian multiplier λ, carry out the variation of $L = \sigma^2 + \lambda b$ with respect to Ψ and thus find the filter Ψ minimizing L. The solution to this constrained minimization in the Fourier domain is given by

$$\hat{\Psi}(\boldsymbol{k}) = \frac{1}{(2\pi)^2}\left[\int_{\mathbb{R}^2} \frac{|\hat{\tau}(\boldsymbol{k})|^2}{P_N(k)}\mathrm{d}^2\boldsymbol{k}\right]^{-1} \frac{\hat{\tau}(\boldsymbol{k})}{P_N(k)}\,, \tag{4.5}$$

where τ is the expected shear profile of a NFW halo and P_N is the noise power spectrum, on which we will focus on Section 4.1.2. The filter is therefore constructed to be most sensitive for those spatial frequencies where the signal $\hat{\tau}$ is large and the noise P_N is low. A back-Fourier transform must be applied to $\hat{\Psi}(k)$ since the convolution to compute the estimator is done in the real domain.

It must be emphasized that when the signal is the tangential shear the estimator 4.2 is mathematically similar to the Aperture mass as defined in Eq. 2.49:

$$A_{est}(\boldsymbol{\theta}) = \int_{\mathbb{R}^2} \gamma_+(\boldsymbol{\theta'};\boldsymbol{\theta})\Psi(\boldsymbol{\theta} - \boldsymbol{\theta'})\mathrm{d}^2\boldsymbol{\theta'}\,. \tag{4.6}$$

Their difference is that the optimal weak lensing filter is not defined from an arbitrary compensated weight function according to Eq. 2.57, but defined so that it maximizes the signal-to-noise ratio and minimizes the contamination induced by the large-scale structure; the properties of the signal-to-noise ratio differ from that of the aperture mass.

4.1.2 Noise properties

It is convenient to describe the statistical properties of the noise contributions in the Fourier space through the *power spectrum* $P_N(k)$:

$$\langle\hat{N}(\boldsymbol{k})\hat{N}^*(\boldsymbol{k'})\rangle = (2\pi)^2\delta_D^{(2)}(\boldsymbol{k'} - \boldsymbol{k})P_N(k). \tag{4.7}$$

It is possible to identify three main noise components:

- Poisson shot noise proportional to the number density of galaxies n_g. This noise arises because the galaxy ellipticities are measured at random discrete positions. It is constant in the Fourier domain (white noise).
- White noise due to intrinsic ellipticities of galaxies. This noise arises because the determination of a single galaxy ellipticity is a very noisy measurement of the shear. Such contribution is proportional to the variance of the intrinsic galaxy ellipticity $\sigma^2_{\epsilon_s}$ and it is independent from the scale k. In first approximation, the background galaxies are randomly positioned and oriented.
- Noise caused by the large-scale structures. When the matter distribution of the LSS is projected along the line of sight, random peaks are contaminating the signal. We model such contribution through the linear dark-matter power spectrum $P_\gamma(k)$, discussed in the introduction of this chapter. Weak lensing by the large-scale structures is well described by an isotropic Gaussian random field.

We can assume these components to be random with zero mean and isotropic such that their statistical properties are independent of the position in the sky. The ellipticity power spectrum P_ϵ models the first two noise components combined, the total noise power spectrum is thus:

$$P_N(k) = P_\epsilon + P_\gamma(k) = \frac{1}{2}\frac{\sigma^2_{\epsilon_s}}{n_g} + \frac{1}{2}P_\kappa(k). \tag{4.8}$$

The cosmic shear power spectrum is equal to the power spectrum of the convergence P_κ generated by large-scale structure, and the factor $1/2$ arises because only one component of the ellipticity (the tangential shear γ_+) contributes to the measurement, thus $P_\gamma = \frac{1}{2}P_\kappa$. This power spectrum is given by (Umetsu, 2020):

$$P_\kappa(k) = \frac{9H_0^4\Omega_m^2}{4c^4}\int_{\chi_{s,i}}^{\chi_s}\frac{\bar{W}(\chi;\chi_s)}{a^2(\chi)}P_\delta\left(\frac{k}{f_K(\chi)};\chi_s\right)\mathrm{d}\chi\ , \tag{4.9}$$

where χ is the comoving distance, χ_s the comoving distance of the sources and $\chi_{s,i}$ is the minimum comoving distance of the sources in the dataset, a is the scale factor of the Universe, $\bar{W}$ is a weight function, P_δ is the three-dimensional matter power spectrum, K is the spatial curvature, $f_K(\chi)$ is the comoving angular diameter distance. As for the weight function $\bar{W}$, it is defined from the (normalized) source distance distribution $G(\chi)$ we have:

$$\bar{W}(\chi;\chi_s) \equiv \int_\chi^{\chi_H} G(\chi_s)\frac{f_K(\chi_s-\chi)}{f_K(\chi_s)}\ , \tag{4.10}$$

where χ_H is the horizon distance defined as the comoving distance obtained for infinite redshift. The comoving distance distribution can be derived from the redshift distribution: $G(\chi)\mathrm{d}\chi = p_z(z)\mathrm{d}z$. In our analysis, we selected a distribution with the functional form $p(z) = \frac{z^a+z^{ab}}{z^b+c}$.

For the computation of this power spectrum, we used the *CosmoBolognalib*, a large set of Open Source `C++` numerical libraries for cosmological calculations (Marulli et al., 2016). The power spectrum is computed given a source redshift, its value represents a typical redshift of the sources, for instance, the mean or median redshift of the source galaxies. The code that computes the cosmic shear power spectrum

inside the *CosmoBolognalib* uses the function `cbl::cosmology::Cosmology::Pk_DM` to compute P_δ, and it assumes a ΛCDM cosmology with the following cosmological parameters: $h = 0.7$ for the normalized Hubble constant and $\Omega_{m,0} = 0.27$, $\Omega_{\Lambda,0} = 0.73$, $\Omega_{bar,0} = 0.046$, $\Omega_{rad,0} = 0.0$ for the density parameters of total matter, cosmological constant, baryons and radiation respectively.

4.1.3 Halo lensing signal model

For the signal required to build the filter we first assume that galaxy clusters are on average axially symmetric thus $\tau(\boldsymbol{\theta}) = \tau(|\boldsymbol{\theta}|)$ and possess a Navarro-Frenk-White (NFW) density profile (Navarro et al. (2004); Navarro (1996)). The profile is given by

$$\rho(r) = \frac{\rho_s}{\left(\frac{r}{r_s}\right)\left(1+\frac{r}{r_s}\right)^2}\,, \tag{4.11}$$

where r_s represents the characteristic scale radius where logarithmic slope $\frac{\mathrm{d}\,\ln\rho}{\mathrm{d}\,\ln r}$ equals -2. The normalization $\rho_s = 4\rho(r_s)$ is a characteristic density. When $r \ll r_s$ the density decrease as $\rho \propto r^{-1}$, when $r \gg r_s$ as $\rho \propto r^{-3}$. The two parameters of the NFW profile are not independent, this profile actually depends only on one parameter. This parameter is the *overdensity mass* M_Δ, defined by integrating Eq. 4.11 up to the *overdensity radius* r_Δ at which the mean internal density is Δ times the critical density of the Universe at halo redshift $\rho_c(z_h)$ as defined in Eq. 1.13. For a NFW halo the *concentration parameter* is defined as $c_\Delta \equiv \frac{r_\Delta}{r_s}$; the typical density is then:

$$\rho_s = \frac{\Delta}{3}\frac{c_\Delta^3}{\ln(1+c_\Delta) - \frac{c_\Delta}{1+c_\Delta}}\rho_c(z_h). \tag{4.12}$$

The NFW profile thus depends only on the concentration parameter c_Δ. By convention we choose Δ to be 200. The gravitational lensing properties of the NFW lens have been widely explored (Meneghetti et al., 2003). Its lensing potential is:

$$\psi(x) = 4\kappa_s h(x), \tag{4.13}$$

with $x = \frac{r}{r_s}$ (adimensional projected distance from the lens center in units of the distance scale),

$$h(x) = \frac{1}{2}\ln^2\left(\frac{x}{2}\right) + \begin{cases} -2\mathrm{arctanh}^2\sqrt{\frac{1-x}{1+x}} & (x<1) \\ 0 & (x=1) \\ 2\mathrm{arctanh}^2\sqrt{\frac{x-1}{x+1}} & (x>1) \end{cases} \tag{4.14}$$

and $\kappa_s = \rho_s r_s \Sigma_{crit}^{-1}$. The shear profile can be computed from the potential by using Eq. 2.18 and Eq. 2.19, the convergence using Eq. 2.8, and the reduced shear profile with Eq. 2.30 (Wright and Brainerd, 2000). The Fourier transform of the reduced shear profile $g(\boldsymbol{x})$,

$$\hat{\tau}(\boldsymbol{k}) = \hat{g}(\boldsymbol{k}) = \int_{\mathbb{R}^2} g(\boldsymbol{x}) e^{i\boldsymbol{x}\cdot\boldsymbol{k}} \mathrm{d}^2\boldsymbol{x}\,, \tag{4.15}$$

is the halo signal necessary to compute the filter. The reduced shear of a NFW halo is shown in real space $g(|\boldsymbol{x}|)$ in the left panel of Figure 4.1, while on the top right panel of Figure 4.2 we have $\hat{g}(|\boldsymbol{k}|)$, the profile in the Fourier space.

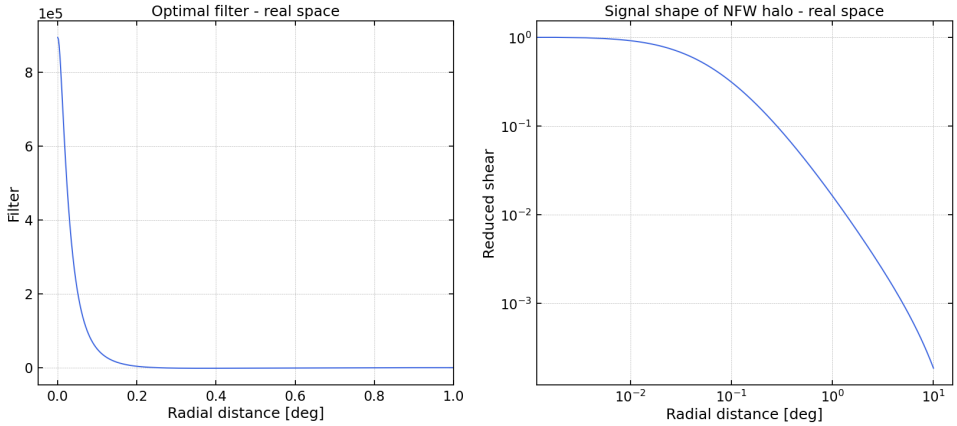


Figure 4.1 *On the left panel we show the optimal filter Ψ and the reduced shear profile from a NFW halo on the right panel, both in real space. The filter is built taking the model of a NFW halo of virial mass 10^{15} $M_{\odot}/h$ at redshift 0.4. The filter in real space is actually obtained from Eq. 4.5 with a Fourier transform. In the* ***left panel*** *it is possible to see the central peak of the optimal filter, responsible for creating signal spots in the output maps where the halos create peaks of lensing signal; in most of the optimal filters we should find outside the peak a negative region, responsible for the throughs of signal, but the optimization applied to the filter we used resolved this problem. In the* ***right panel*** *we show the signal profile τ of the NFW halo equivalent to the reduced shear $g(\boldsymbol{x})$.*

We presented all the components that participate in the construction of the optimal filter. We can finally look at the profiles: the left panels of Figure 4.1 and 4.2 show the optimal filter in the real and the Fourier space. In real space, it shows a central peak, responsible for creating signal spots in the output maps where there are halos, and a negative region outside the peak, responsible for creating troughs with lower signals outside the spots. In a typical weak-lensing amplitude map (Figure 4.3), thus when the signal processed by the filter is the tangential shear, the typical signature of a halo detection is a spot with high signal-to-noise ratio SNR. In the bottom panel of 4.2 we have the optimal profile of the filter in the Fourier space. We can see that the filter has a peak around those spatial frequencies with a high ratio between the reduced shear of the NFW halo $\hat{\tau}$ and the noise power spectrum P_N as described in 4.5.

4.2 Augmenting AMICO for weak-gravitational lensing measures

In the first part of Chapter 4 we illustrated the advantages of matched filter techniques and in particular of AMICO: the code can be applied to any catalog of galaxies describing arbitrary properties so that the routines and the classes of the library, on which the code is based on, can be adopted for different data sets and observables.
We implemented a new weak-lensing branch in AMICO by exploiting the polymorphism of `C++` to minimize the modifications of the well-tested existing routines. AMICO-WL follows the layout of the optical routine that can be broken down into

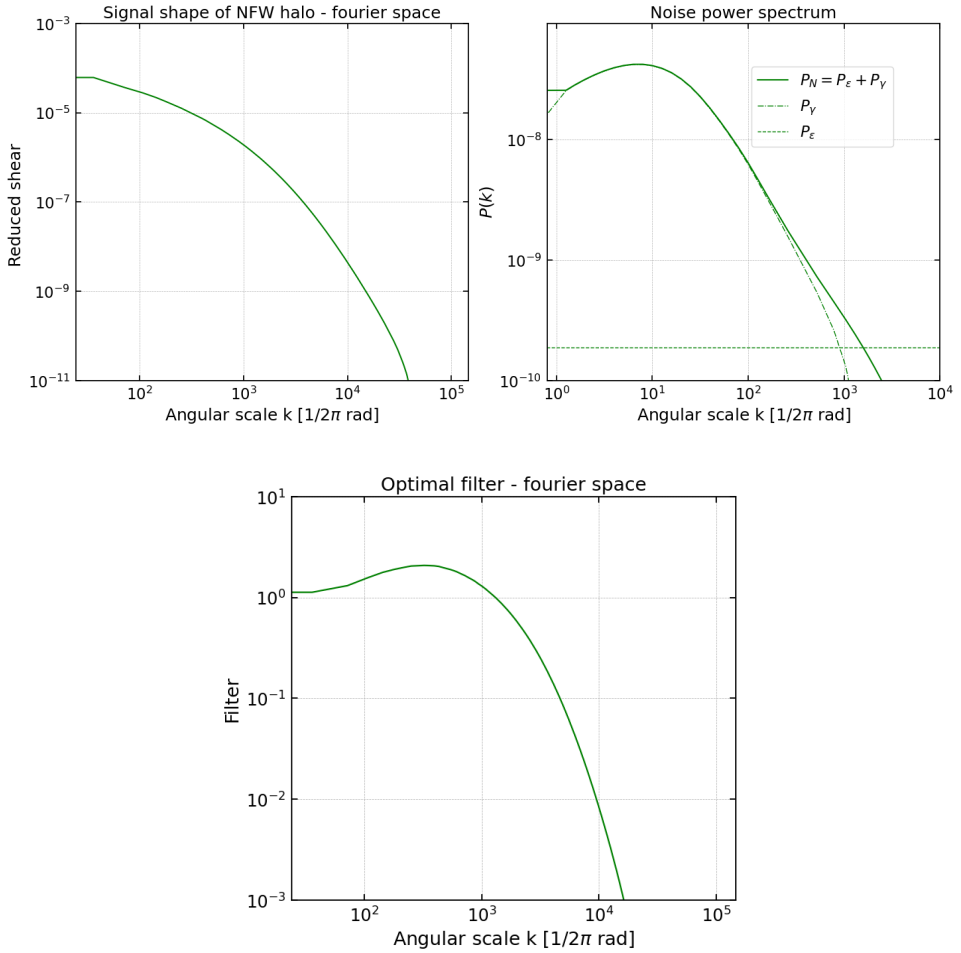


Figure 4.2 *We show the reduced shear profile from a NFW halo $\hat{\tau}$ on the top-left panel, on the top-right panel the power spectrum of the noise $P_N(k)$ and on the bottom the optimal filter $\hat{\Psi}$. The filter is built taking the model of an NFW halo of virial mass 10^{15} $M_\odot/h$ at redshift 0.4. The background sources are assumed at redshift $z = 1.0$, used to compute the noise power spectrum of the large-scale structure P_γ. The galaxy white noise is $P_\epsilon = 1.88 \times 10^{-10}$. These quantities allow us to compute the filter in the Fourier space $\hat{\Psi}$ we have in* ***bottom panel****, where we can see that the filter has a peak around those spatial frequencies with a high ratio between the reduced shear of the NFW halo $\hat{\tau}$ and the noise power spectrum P_N as described in 4.5. In the* ***top-right panel*** *we show P_N with the solid line that is composed of the white noise of galaxies (shot noise and intrinsic ellipticity noise) P_ϵ and the power spectrum of large-scale structure P_γ, the dashed lines. In the* ***top-left panel*** *we have $\hat{\tau}$ was obtained with a Fourier transform from the NFW profile in real space (right Fig. 4.1).*

three main steps:

1. **Initialization**. The code sets up the environment for the application of the algorithms reading instructions from the input files (e.g. survey's galaxy catalog).
2. **Map creation**. The optimal filter technique is applied and both the amplitude and variance maps are produced.
3. **Detection**. The SNR map is computed and the peaks are detected through the *cleaning* procedure.

After discussing briefly the necessary input of the program, in Sections 4.2.2 and 4.2.3 we are going into the details of the maps creation and the detection procedure explaining the implementation of the weak-lensing data analysis.

4.2.1 Input of AMICO-WL

The first and fundamental file that AMICO-WL needs is the data set, the catalog of galaxies with all properties measured that comes from surveys or simulations. In the cluster detection through weak-gravitational lensing, the properties that AMICO needs for each galaxy are: two sky coordinates of the galaxy, they could be cartesian coordinates or sky coordinates (RA, Dec), and the two components of the ellipticity g_1 and g_2 (i.e. the expected values of the reduced shear). These four quantities are mandatory for the detection but other additional properties can be read as well, for example, the redshift of the galaxy z_s and the weak lensing weight w of the galaxy, generally defined as the inverse-variance weight of the ellipticity.

The second input file is the initialization file which contains all the parameters necessary for the execution of AMICO-WL. This is passed through a command line argument to the executable, every line has a standard layout for the input parameters with format (`parameter = value`).

The third and last input file contains the optimal filter Ψ in the real domain. The file must contain three columns: (1) one for the radial distances r in *degrees* at which the filter is evaluated, (2) one for the value of the filter $\Psi(r)$ evaluated at the radial distance r. The radial values must be sampled regularly and linearly to ensure a correct interpolation of the filter function. Finally, (3) a third column must contain the value of the reduced shear of the chosen NFW halo $g(r)$ at the radial distance r. This column is used during the cleaning procedure that will be discussed in Section 4.2.3. The filter file is produced separately with a code that implements Eq. 4.5 and its Fourier transform. This program requires the following input parameters:

- File with the cosmological parameters to set the fiducial cosmological parameters.
- Mass of the fiducial NFW dark matter halo used as a template in units of $\mathrm{M}_{\odot}h^{-1}$.
- Redshift of the typical halo we want to detect.
- Average redshift of the sources.
- White noise amplitude: the contribution of the power spectrum given by the galaxies, defined as P_ϵ in Section 4.1.2.

- Instrumental beam FWHM: the resolution of the filter definition computed as the inverse of the surface number density of galaxies.
- File containing the cosmic shear power spectrum of the convergence P_γ to be used for the noise definition.
- Optional: scale radius to set the angular size of the template halo. This overrides the size set through the choice of the halo mass.

The following Sections 4.2.2 and 4.2.3 discuss the execution of AMICO-WL and explain the creation of the signal maps and the cleaning procedure to identify the detections.

4.2.2 Creation of the weak-lensing signal maps

For the creation of the signal maps, the algorithm must identify all galaxies falling in each pixel: in order to speed up this operation, galaxies are first sorted in a linked list.

The contribution of each galaxy on all pixels is then summed to the amplitude map accounting for the weight provided by the filter. At the same time, the variance signal is evaluated.

The estimator 4.2 is computed by using the tangential component of the ellipticity ϵ_+ (which is a good estimator of the tangential shear γ_+) as

$$E_{map}(\boldsymbol{\theta}) \equiv A_{est}(\boldsymbol{\theta}) = \frac{1}{n_g} \sum_k \epsilon_+(\boldsymbol{\theta};\boldsymbol{\theta_k}) w_k \Psi(|\boldsymbol{\theta_k} - \boldsymbol{\theta}|) \ , \tag{4.16}$$

where the sum is extended to all galaxies within a circular area centered on $\vec{\theta}$ and extending up to a *cut radius* which is a user defined parameter. If the lensing weights w_k are not used, they are all set equal to the default value: 1. The normalization factor n_g is the effective number density of galaxies given by the sum of their weights. The tangential shear is computed by using Eq. 2.60. Eq. 4.16 implements the aperture mass estimator defined in Eq. 2.70 when the filter function is the optimal linear filter in the real domain. Along the amplitude map, the code produces the variance map, i.e. the map noise estimate for the variance obtained as $\sigma^2 = \langle (A_{est} - A)^2 \rangle$:

$$\sigma^2_{E_{mode}}(\boldsymbol{\theta}) \equiv \sigma^2_{A_{est}}(\boldsymbol{\theta}) = \frac{1}{2n_g^2} \sum_k |\epsilon_+(\boldsymbol{\theta};\boldsymbol{\theta_k})|^2 w_k^2 \Psi^2(|\boldsymbol{\theta_k} - \boldsymbol{\theta}|) \ . \tag{4.17}$$

This noise estimate is only due to galaxy noise. To get the complete noise (with the LSS component) we must add the constant:

$$C_{LSS} = \frac{1}{2\pi} \int_0^\infty |\hat{\Psi}(\boldsymbol{k})|^2 P_\gamma(k) k \mathrm{d}k \ . \tag{4.18}$$

The sum of Eqs. 4.17 and 4.18 is the actual map of the noise variance. Since the large-scale structure contributes to the variance with a constant, it is applied to all pixels at the end of the process.

The same algoritmic infrastructure is also used to evaluate the B-mode of the lensing signal which is given by

$$B_{map}(\boldsymbol{\theta}) = \frac{1}{n_g} \sum_k \gamma_\times(\boldsymbol{\theta};\boldsymbol{\theta_k}) w_k \Psi(|\boldsymbol{\theta_k} - \boldsymbol{\theta}) \tag{4.19}$$

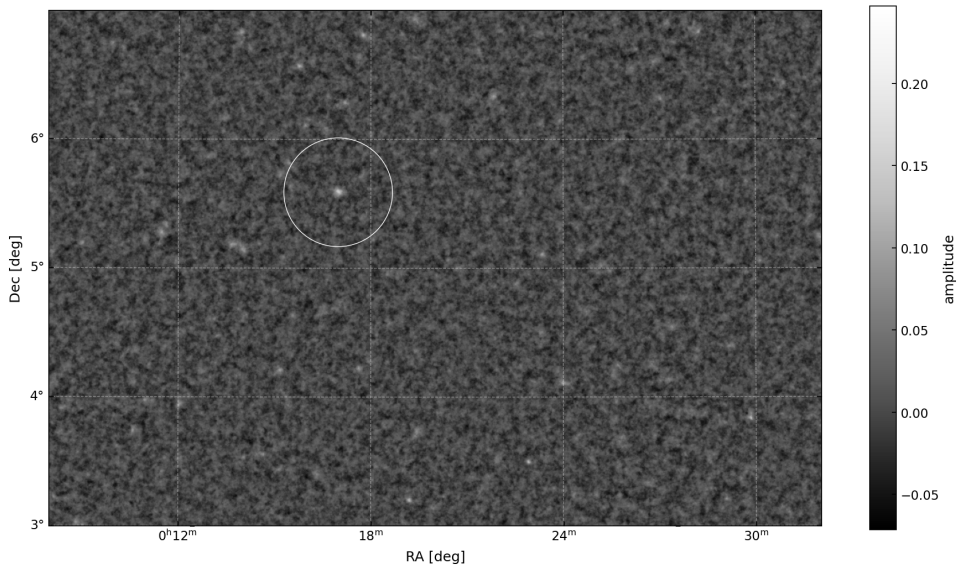


Figure 4.3 *Part of an output map from a set of simulated data we are going to describe in Chapter 5. The map is related to the E-mode of the shear, thus the true lensing signal. It is possible to catch several peaks of the signal, one of the major ones is selected inside the white circle and represents a large cluster.*

to produce the *B-mode amplitude map*. The only difference with respect to Eq. 4.16 is the use of the cross shear $\gamma_\times$ defined in Eq. 2.61. Such B-mode can be used to estimate the noise fluctuations as they do not contain any lensing contribution by construction. The corresponding variance map of the B mode can be obtained by:

$$\sigma^2_{B_{map}}(\boldsymbol{\theta}) = \frac{1}{2n_g^2} \sum_k |\gamma_\times(\boldsymbol{\theta}; \boldsymbol{\theta_k})|^2 w_k^2 \Psi^2(|\boldsymbol{\theta_k} - \boldsymbol{\theta}|) \ , \tag{4.20}$$

to which Eq. 4.18 must be added to include the LSS contribution.

4.2.3 Detection algorithm and cleaning procedure for the weak-lensing case

The halos with the largest signal are visible as clear peaks surrounded by a circular region with values that are slightly smaller than the background, as it can be seen in Figure 4.3. This is the typical signature of the weak lensing filter radial dependency, which is characterized by a central peak (that produces the maxima) and a slightly negative region just outside the peak (responsible for the minima). The values in the amplitude map can either be positive or negative because of the noise and because the cosmic shear signal is also characterized by voids leading to "depressions" in the signal.

Once the amplitude and variance maps have been created, the code identifies the peaks. The detection is not carried out on the amplitude map but on the signal-to-noise ratio which has to be computed first. The algorithm identifies the "relevant pixels", i.e. those pixels whose signal-to-noise ratio value is above a certain threshold provided by the user.

The procedure then identifies the maximum which is marked as the first detection to be stored together with its properties: the ID in order of detection (order of SNR), the X and Y pixel position, the physical position in RA and Dec, the SNR and the value of the amplitude map. Then the algorithm removes the signal contribution of this detection from the amplitude map.

At the same time, the depressions caused by the filtering procedure that surrounds the detection are also removed. The practical procedure to implement this cleaning involves the creation of a *dummy galaxy* for each real galaxy in the catalog with the same position but with the two components of the shear effect produced by the modeled lensing signal associated with the detection that is then removed from the amplitude map: in this way, the map is "cleaned" from all lensing effect of the detections.

This is done by removing the filtered lensing signal modeled with the same template adopted for the filter construction. Here, the shear modulus is modeled at the position $\boldsymbol{\theta}_d$ as

$$|\gamma_d(\boldsymbol{\theta};\boldsymbol{\theta}_d)| = \mathcal{A}\cdot\tau(\boldsymbol{\theta};\boldsymbol{\theta}_d)\ , \tag{4.21}$$

with a shear profile, $\tau(\boldsymbol{\theta})$, normalized by the detection amplitude, $\mathcal{A}$. We can write the complex shear attributed to the detection as:

$$\gamma_d = \gamma_{d,1} + i\gamma_{d,2} = |\gamma_d|(\cos 2\varphi + i\sin 2\varphi)\ . \tag{4.22}$$

In Section 2.2.1 we have shown that the shear effect given by gravitational lensing exists only in the γ_+ components, thus the shear is tangential or radial with respect to the line connecting the center of the detection (assumed radially symmetric) and the considered sky position centered on a galaxy, see Figure 4.4, which shows the two possible geometrical configurations of the galaxy shear.

This simple geometric consideration sets the two components of the tangential shear:

$$\begin{aligned}\gamma_{d,1}(\boldsymbol{\theta};\Delta\theta) &= |\gamma_d(\boldsymbol{\theta};\Delta\theta)|\cdot\cos 2\varphi = -|\gamma_d(\boldsymbol{\theta};\Delta\theta)|\cdot\cos 2\phi \\ \gamma_{d,2}(\boldsymbol{\theta};\Delta\theta) &= |\gamma_d(\boldsymbol{\theta};\Delta\theta)|\cdot\sin 2\varphi = -|\gamma_d(\boldsymbol{\theta};\Delta\theta)|\cdot\sin 2\phi\ .\end{aligned} \tag{4.23}$$

Since on small angles, the flat sky approximation is valid, the galaxy position with respect to the detection center can be safely expressed in a cartesian frame as $\Delta\theta = \theta - \theta_d = \Delta\theta_1 + i\Delta\theta_2$ and the shear can be expressed in two ways, explicit functions of the cartesian components and using the trigonometric equations:

$$\begin{aligned}\gamma_{d,1}(\boldsymbol{\theta};\Delta\theta) &= -|\gamma_d(\boldsymbol{\theta};\Delta\theta)|\cdot\frac{\Delta\theta_1^2-\Delta\theta_2^2}{\Delta\theta_1^2+\Delta\theta_2^2} = -|\gamma_d(\boldsymbol{\theta};\Delta\theta)|\cdot\cos\left(2\arctan\frac{\Delta\theta_1}{\Delta\theta_2}\right) \\ \gamma_{d,1}(\boldsymbol{\theta};\Delta\theta) &= -|\gamma_d(\boldsymbol{\theta};\Delta\theta)|\cdot 2\frac{\Delta\theta_1\Delta\theta_2}{\Delta\theta_1^2+\Delta\theta_2^2} = -|\gamma_d(\boldsymbol{\theta};\Delta\theta)|\cdot\sin\left(2\arctan\frac{\Delta\theta_1}{\Delta\theta_2}\right)\end{aligned} \tag{4.24}$$

The routine implements the trigonometric expressions (last passage of Eq. 4.24) that, thanks to the optimized *sin*, *cos* and *arctan* functions in `C++`, are more efficient from a computational point of view.

After the removal of the detection contribution, the SNR is re-evaluated and the algorithm continues by considering the next maximum in the relevant pixels

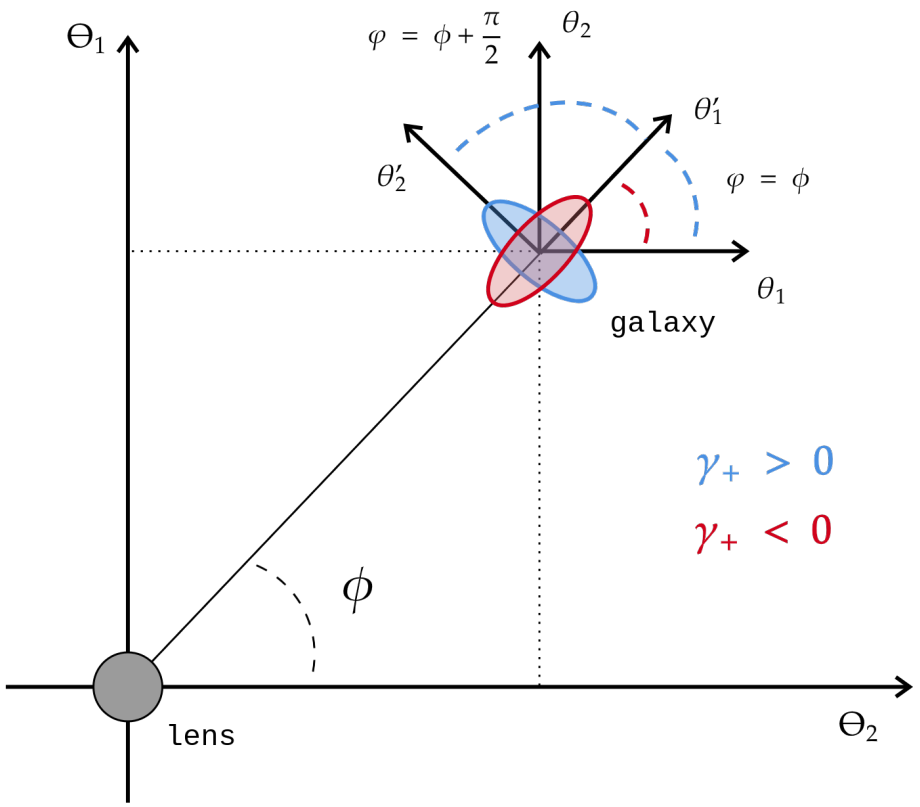
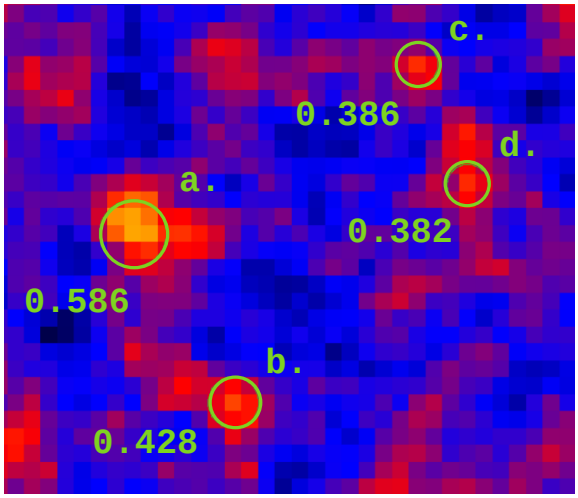


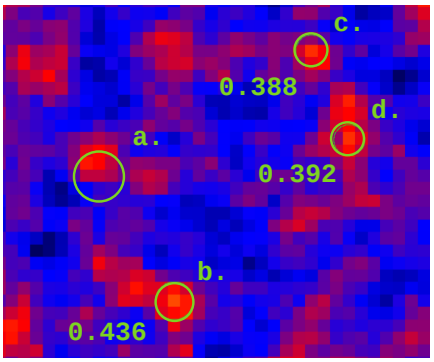
Figure 4.4 *The diagram represents the geometrical configuration of the distortion applied on a background galaxy from a gravitational lens. For the sake of clarity we use fake proportions of the dimensions. We show the system of reference comoving with the lens* (Θ_1, Θ_2) *and comoving with the galaxy* (θ_1, θ_2)*. The system* (θ_1', θ_2') *is the coordinate system obtained from a rotation* $R(\phi)$ *that diagonalizes the lensing jacobian matrix, that is true because* ϕ *is the angle relative to the connecting line between lens and galaxy. The shear components in this coordinates system yield the tangential and cross shear components Eq. 2.63). We have then* $\gamma_+ > 0$ *when the distortion is perfectly tangential (blue galaxy) and the shear angle is* $\varphi = \phi$*, and* $\gamma_+ < 0$ *when the distortion is perfectly radial (red galaxy) and the shear angle is* $\varphi = \phi + \frac{\pi}{2}$*.*

reiterating the process. It may happen that after the cleaning, some peaks that initially had a SNR value below the detection threshold can be boosted such to be detectable. An example is given in Figure 4.5 where it is shown the cleaning procedure applied to the detection of 4 objects in a region of the map: following the sequence of detection/cleaning it is possible to see the change in the amplitude values of the peaks. Therefore the detections are not identified with a perfect monotonic decrease in their value of SNR.

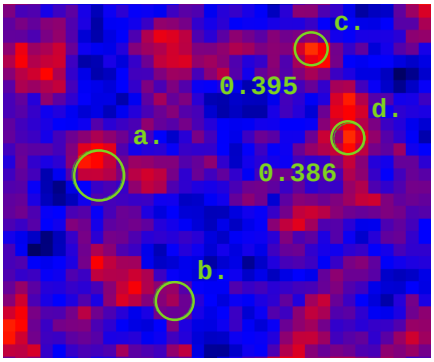
At the end of the procedure, the algorithm reaches the minimum signal-to-noise ratio threshold fixed by the user and the iteration stops. With the catalog of detections, the code outputs the amplitude map removed of all the lensing signals from the dark-matter halo candidates. This map is useful to test the results as it should resemble the B-mode map where no lensing signal, apart from a uniform contribution of the LSS, is present.



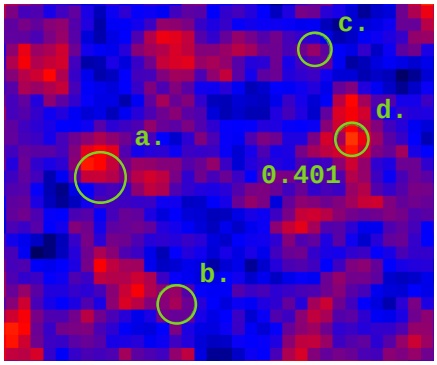
(i) *pre-cleaning*



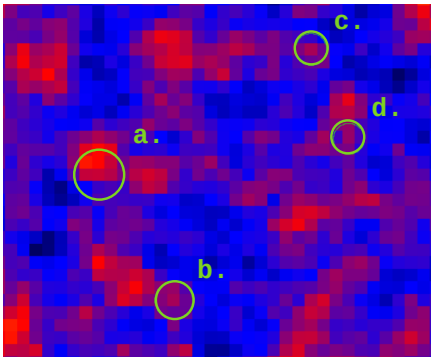
(ii) *cleaning of a*



(iii) *cleaning of b*



(iv) *cleaning of c*



(v) *cleaning of d*

Figure 4.5 *Example of the cleaning procedure on a small region with 4 signal peaks. The circles in the figures are centered in the pixel with the maximum amplitude value reported on the bottom left, while the identifying letter is on the top right. The map before any cleaning (Fig. 4.5i) shows 4 peaks with maximum amplitudes of 0586 (a), 0.428 (b), 0.386 (c), and 0.382 (d). The algorithm thus finds first (a) and removes the expected shear effect from the map (Fig. 4.5ii), the other peaks find themself with an increased amplitude (b, c, d with respectively 0.436, 0.388, and 0.392). Then it is the time of (b) to be detected and cleaned (Fig. 4.5iii), increasing peaks (c) and (d) to 0.395 and 0.386 respectively. In Fig. 4.5iv we have the map after the third peak (c) is detected and removed, at this point (d) has increased its amplitude again to 0.401 and it is finally detected and removed leaving this region of the map as pure background (Fig. 4.5v).*

Chapter 5

Analysis of the Euclid simulations with AMICO-WL

In this chapter, we describe the results of the application of the AMICO-WL algorithm, discussed in Chapter 4, to a simulated catalog of galaxies ellipticities lensed by dark matter halos and by large-scale cosmic structure. The simulation, see Section 5.2.1 reproduce the characteristics of the data expected for the *Euclid* mission.

Section 5.1 will describe Euclid and its scientific goals; Section 5.2.1 will provide a description of the data sets used to test AMICO-WL; in Section 5.3 we will analyze the application of AMICO-WL to the simulated data; in Section 5.4 it is described the matching procedure applied to test the code; in Section 5.5 we will study the results of the matching in terms of purity and completeness of the sample; in the final Section 5.6 we will compare the results of two simulated datasets.

5.1 The Euclid mission

Euclid (Laureijs et al. (2011);Scaramella et al. (2022)) is a visible-to-near-infrared space telescope currently under development by the European Space Agency (ESA) and the Euclid Consortium; the launch is scheduled to occur in July 2023, see the EUCLID website [1]. The goal of Euclid mission is in fact, to better characterize dark energy and dark matter, which make up most of the energy content of the Universe, as seen in Chapter 1. For that purpose, the mission will investigate the distance-redshift relationship and the evolution of cosmic structures by measuring shapes (i.e. ellipticities) and redshifts of galaxies and clusters of galaxies out to redshifts ~ 2 (a look-back time of ≈ 10 Gyr). In this way, Euclid will cover the entire period over which dark energy played a significant role in accelerating the expansion of the Universe. Some main topics that Euclid is going to help investigate are:

- Dark matter distribution in the Universe
- History of the expansion of the Universe
- The nature of dark energy and the evolution of its equation of state

[1] https://www.esa.int/Science_Exploration/Space_Science/Euclid

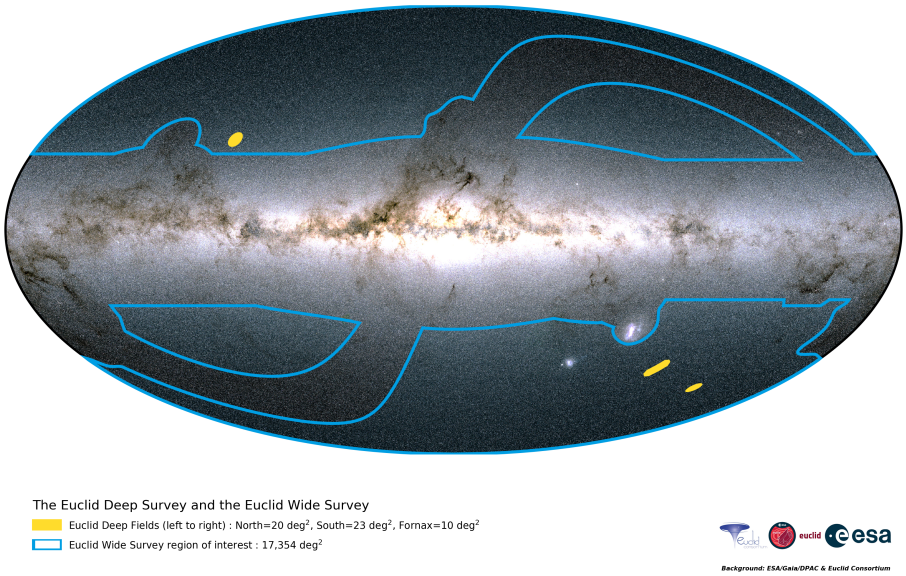


Figure 5.1 *We present the sky coverage of Euclid surveys. Inside the blue lines, we have the field of view (FOV) of the Wide Survey of 17 354 deg^2 while the three yellow patches represent the FOVs of the Euclid Deep Surveys which coverage goes from 10 to 23 deg^2. Credits: ESA/Euclid Consortium. Acknowledgment: Euclid Consortium Survey Group*

- Large-scale structure formation in the Universe

Euclid is optimized for two primary cosmological probes:

- **Weak gravitational Lensing** by measuring distortions of galaxy images by mass inhomogeneities along the line-of-sight to map the dark matter distribution and determine the impact of dark energy.

- **Baryonic Acoustic Oscillations**, wiggle patterns imprinted in the spectroscopic clustering of galaxies on large scales, which provide a standard ruler to measure the expansion of the Universe and the acceleration caused by dark energy.

Euclid will survey the sky in a "step and stare" mode: the telescope will point to a position in the sky and perform imaging and spectroscopic measurements on an area of ~ 0.5 deg^2 around this position. The visual imager is characterized by a sensitivity of 25 mag for the visual imager and 24 mag for the near-infrared photometer. The sky coverage is driven by the wide-survey requirement to cover 15,000 deg^2 of the extragalactic sky (more than 35% of the celestial sphere) during the mission lifetime of 6 years. Additionally, a deep survey will be performed in three deep fields (40 deg^2 in total), reaching two magnitudes deeper than the wide survey. In Figure 5.1 we show the Euclid fields of view, both wide and deep.

5.2 Simulated Euclid catalogs

The catalog of ellipticities onto which we applied the code derives from a parent cosmological simulation produced by Carlo Giocoli, while the noise in the catalog of ellipticities was added by Sandrine Pires. These simulated data were created as a part of a "test challenge" organized inside the *Euclid SWG Clusters of Galaxies* and more specifically as a part of the activities of WP 10, "Weak Lensing Selected Clusters". The idea is to evaluate the performance of different methods for detecting galaxy clusters through weak lensing that could be implemented alongside the ones based on photometric information, which are already implemented in the official Euclid data analysis pipeline. These algorithms, selected through a similar challenge (Adam et al., 2019), are AMICO (Bellagamba et al., 2018) and the *PZWav* (Gonzalez, 2019), code based in an adaptive wavelet approach. We are now going into the details of the simulated ellipticity catalog and the parent cosmological simulation from which it has been extracted.

5.2.1 The numerical simulations

The cosmological simulation is the result of a dark-matter-only N-body run, carried out with `GADGET-2` (Springel (2005); Springel et al. (2001b)). The simulation assumes a flat ΛCDM model with normalized Hubble constant $h = 0.67$, dark matter density parameter $\Omega_{CDM} = 0.27$, baryon density parameter $\Omega_{bar} = 0.05$, amplitude of the density fluctuations $A_s = 2.1265 \times 10^{-9}$ (used instead of σ_8, the r.m.s. density fluctuations in spheres of 8 h^{-1}Mpc as in Sect. 3.1), spectral index of the power spectrum of the initial density fluctuations $n_s = 0.96$.

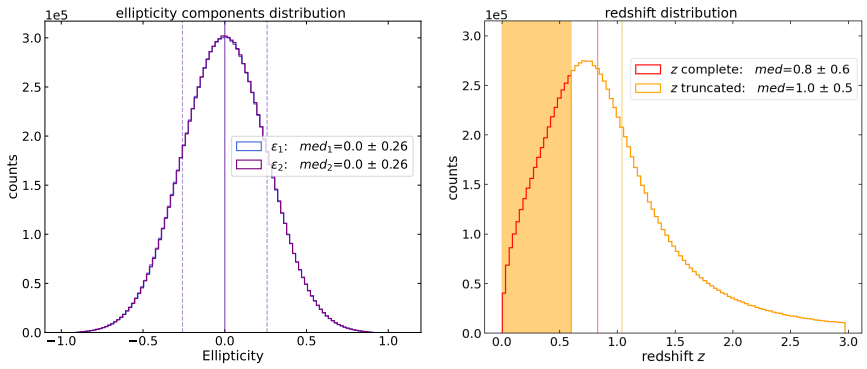
Initial conditions have been created at redshift $z = 99$ using `N-GenIC` and assuming the corresponding power spectra rescaled at that redshift using `CAMB`. It follows the evolution until $z = 0$ of 1024^3 collisionless dark matter particles in a periodic cube with 1 h^{-1}Gpc on a side. At each snapshot, halos and subhaloes have been identified sing `SUBFIND` (Springel et al., 2001a). Simulating shear measurements, i.e. ellipticities in the galaxies, means being able to reproduce and investigate the propagation of light through the inhomogeneous universe. In order to do so, several snapshots of the simulation – at different redshifts have been piled together to construct simulated past-light cones for the ray-tracing analysis. Those have been built using the `MapSim` routines :`iMapSim` and `ray-MapSim` (Giocoli et al. 2018, 2015); that have been validated by comparison with other algorithms (Hilbert et al., 2020). `iMapSim` requires to set the desired field of view, chosen to be a light cone with a 10 deg aperture, and the number of snapshots ahead of time; the authors saved 43 snapshots from redshift 4 to 0. The value of redshift $z = 4$ was chosen to investigate where the dynamical evolution of the matter-energy components of the Universe starts to leave a mark in the weak lensing observables (Giocoli et al., 2015). Each snapshot consists of a cubic volume containing one realization of the matter distribution at a given redshift. The snapshots are obtained from the same initial conditions, thus, the volumes contain the same cosmic structures at different stages of their evolution, approximately at the same positions in each box. It is necessary to avoid repetitions of the same cosmic structures along one line of sight, thus the snapshots need to be randomized as they cannot be stacked as a consecutive sequence. The randomization is described in (Roncarelli et al., 2007). Then the snapshots are converted from 3-dimensional volumes into 2-dimensional mass distribution by projecting the particle positions to the nearest pre-determined plane, maintaining

their angular positions. Finally, the ray-tracing, in Born approximation, is performed using `ray-MapSim` considering the 43 planes defined before as lens planes. Lens equations are solved on the two-dimensional grid using the `MOKA` (Giocoli et al., 2012) libraries, which allows going from the convergence to the corresponding shear maps. The light cone has been populated with sources assuming random right ascension and declination, with a redshift distribution that resembles the one expected for the Euclid wide-field survey, peaked at $z \sim 1$. The analyses done in this work use the shear catalogs produced by these simulations, which also provide the distribution of intrinsic ellipticity for the galaxies.

5.2.2 The simulated data

The test of AMICO-WL has been performed on two simulated catalogs of shear measurements, i.e. ellipticity components. The first simulated catalog consists of around 10 million galaxy positions with ellipticities and redshift, randomly distributed in one field of 10 deg $\times$ 10 deg, with right ascension and declination both in the range [0.0,10.0] deg. The second catalog is a subset of the first obtained by cutting out all the galaxies with a redshift smaller than 0.6. This is done to attempt a cleaning of the data set from the noise that foreground galaxy ellipticities produce on the weak lensing measures. In fact, since we expect to detect galaxy clusters mainly below $z \approx 0.6$, by cutting the data below that value it is possible to select those galaxies that are almost exclusively background galaxies, and thus lensed sources, for the dark-matter halos we are trying to detect. The galaxies at redshifts below 0.6 can be background sources for some clusters but also foreground galaxies for others and thus noise for the constraining of the lensing signal. Therefore with the second catalog of galaxies, we expect to find more halos and increase the signal-to-noise ratio.
In Figure 5.2 we show the dstributions of the main properties of the galaxies: redshift and ellipticity. The distributions of the two ellipticity components are Gaussian with 0 mean and standard deviation 0.26 (see Fig. 5.2i). The redshift distribution is in Fig. 5.2ii, galaxies are within a range of [0,3] with a median redshift of 0.8. One of the main differences between the simulated data and what we expect from Euclid is the fact that in the future real data, we will have galaxies up until $z \sim 2$. We introduced in Section 4.2.2 that the noise component related to the large-scale structure is computed through an external algorithm that we use before launching the AMICO-WL algorithm. Here we present the value of the noise constant from the LSS lensing effect, C_{LSS}, and other properties related to the catalogs, computed as additional outputs of the code:

- **LSS noise component** C_{LSS}: 3.9×10^{-4};
- **galxy number surface density**: 30.0 arcmin^{-2};
- Total number of galaxies: 10 801 552;
- Average galaxy separation on the sky: $0.182'$;
- Mean shear modulus: 0.33;
- Root mean square of the shear modulus: 0.36;
- White noise from galaxies (see Eq. 4.8): $P_\epsilon = 1.88 \times 10^{-10}\text{rad}^2$.



(i) ϵ_1 *and* ϵ_2 *distributions* **(ii)** z *distribution*

Figure 5.2 *Distributions of ellipticities and redshift of the simulated sources in the "Euclid test challenge" shear catalog. In the* ***left panel*** *we show the distribution of the two ellipticities components (in blue and purple) that are almost equivalent being both simulated with mean 0 (vertical line) and standard deviation 0.26 (dashed vertical lines). In* ***right panel*** *the redshift distribution of the simulated galaxies is shown. With the vertical lines, we show the median 0.8±0.6 of the complete catalog (red) and the median 1.0±0.5 of the truncated catalog (orange). We show with an orange band the range of galaxies cut out in the truncated catalog.*

The second catalog is composed of about 7 million galaxies, in the same FOV; the distributions of ellipticities components are equivalent to the ones of the complete catalog (left panel of Fig. 5.2); in the right panel of Figure 5.2 it is visible the clean cut on $z = 0.6$ and the consequent increase in the redshift median. The other interesting properties of this second catalog:

- **LSS noise component** C_{LSS}: 6.1×10^{-4}
- **galxy number surface density**: 20.0 arcmin^{-2};
- Total number of galaxies: 7 449 910;
- Average galaxy separation on the sky: $0.220'$;
- Mean shear modulus: 0.33;
- Root mean square of the shear modulus: 0.36;
- White noise from galaxies $P_\epsilon = 2.74 \times 10^{-10}\text{rad}^2$.

It is interesting to notice that even though the noise of the signal in the E-mode map should decrease, as we will see in the next sections, by cleaning the catalog from the foreground galaxies, the white noise attributed to the distribution of intrinsic galaxy ellipticity increases; we defined this noise component in Eq. 4.8 as $P_\epsilon = \frac{1}{2}\frac{\sigma^2_{\epsilon_s}}{n_g}$, thus as the number surface density of galaxies decreases and the standard deviation of the intrinsic ellipticity stays the same, P_ϵ will increase.

We shall refer to the original catalog as the *complete catalog* and to the redshift cut catalog as the *truncated catalog*. The summary of the two catalogs' properties can be found in table 5.1.

		complete	truncated
N	$[10^6]$	~ 10.8	~ 7.4
σ_ϵ		0.26	0.26
n_g	[arcmin^{-2}]	30.0	20.0
z	[range]	[0.0,3.0]	[0.6,3.0]
$\bar{\Delta}_g$	[arcmin]	0.182	0.220
$\bar{\lvert\gamma\rvert}$		0.33	0.33
$\lvert\gamma\rvert$ r.m.s.		0.36	0.36
P_ϵ	[10^{-10}rad^2]	1.88	2.74
C_{LSS}	$[10^{-4}]$	3.9	6.1

Table 5.1 *Summary table of the main properties of the simulated ellipticity catalogs. In the table, we show the number of galaxies N, the standard deviation of the ellipticity σ_ϵ, the surface number density of the galaxies n_g, the redshift z range, the average separation of the galaxies in the sky $\bar{\Delta}_g$, the mean and r.m.s. of the shear modulus $|\gamma|$, the white noise contributed by the galaxies P_ϵ (shot noise and intrinsic ellipticity noise) and finally the noise constant related to the lensing effect of large-scale structure C_{LSS}.*

From the construction of the simulation described in Section 5.2.1 we have the catalogs of dark matter halos that acted as gravitational lenses for the ray-tracing algorithm. In Section 5.4 we will use this catalog as a reference for the matching procedure, in order to test the reliability of the code. We will need to apply a selection on the structures, in fact, they are given with a set of properties (mass, virial mass, virial radius, position, redshift, comoving distance) that cover large ranges of values, far from the values expected from a dark matter halo associated to a cluster of galaxies. For example the virial mass (M_{200}) goes from a minimum of $4.13 \times 10^{11}\ h^{-1}M_\odot$ to a maximum of $1.39 \times 10^{15}\ h^{-1}M_\odot$.

5.3 Application of AMICO-WL to the mock catalogs

We applied AMICO to the two catalogs introduced in Section 5.2.2, one including all galaxies and one with all galaxies above $z > 0.6$ only. For each data set we optimized the filter according to its specific properties. At this end the cosmic shear power spectrum and the effective density of galaxies have to be adapted. Table 5.1 summarizes the properties of the two filters.

For the template halo we assume a standard set of parameters, used for instance in Maturi et al. (2007), meant to detect high-mass halos that have an intermediate distance between observer and sources, a condition for which the lensing effect is maximized because of the geometrical dependencies of the lensing strength (Pace

		complete	truncated
lens mass	M_l $[h^{-1}M_\odot]$	1×10^{15}	1×10^{15}
lens redshift	z_l	0.4	0.6
source redshift	z_s	1.0	1.0
white noise	N_ϵ	1.27×10^{-10}	1.81×10^{-10}
beam FWHM	b_{FWHM} [arcmin]	0.55	0.65

Table 5.2 *Summary table of the parameters used to compute the filters. We show the mass of the typical lens M_l (i.e. the expected dark matter halo), the redshift of the typical lens z_l, the redshift of the typical source z_s (i.e. the background galaxies), the white noise of the catalog of galaxies N_ϵ and the beam FWHM b_{FWHM}.*

et al., 2007). Figure 5.3 shows the profile of the filters in real space and normalized for the maximum value to better compare them. It is clear that the two filters are almost identical, the parameters that define mainly the shape of the filters, i.e. lens properties, are the same. More importantly, their value is very small at radial distances larger than 0.3 deg. This fact allows to impose a cut-off at that radius reducing the computational costs.

A zoom in of the resulting E-mode and B-mode maps (see Eq. 4.16 and Eq. 4.19) based on the truncated catalog is shown in Figure 5.4. Several peaks are clearly visible in the E-mode map while the B-mode map appears more homogeneous and with a similar distribution of the absolute values of the positive and negative peaks.

In the next sections, we are focusing on the analysis performed on the truncated catalog, the one that we expect to produce the best results, leaving the essential of the analysis of the complete catalog to the Appendix A. In Section 5.6 we will compare the performance of the code with the truncated catalog with the results obtained with the complete catalog.

5.3.1 Criteria for the choice of the detection threshold

We tested four different methods to establish the SNR threshold to be used in the sample selection. Two of them rely on the analysis of the signal-to-noise distribution, the other two instead are based on the statistics of the maxima and/or minima of the maps.

The distribution of values of the E-mode and of the SNR maps (see top panel of Figure 5.5) is a nearly symmetric distribution centered around 0 (if there are no systematic errors). The tail extending toward positive high values is due to the contribution of halos to the lensing signal. In contrast the B-mode map is perfectly symmetric with respect to zero as it contains only the noise fluctuations (shot noise and scatter due to the intrinsic ellipticity of galaxies) and no lensing signal. Thus, the B-mode map provides the statistics of the amplitude values associated to the noise only.

Pixels $E - B$ method: based on these considerations we compute the ratio between the distributions of the pixels SNR of the E- and B-mode maps (bottom

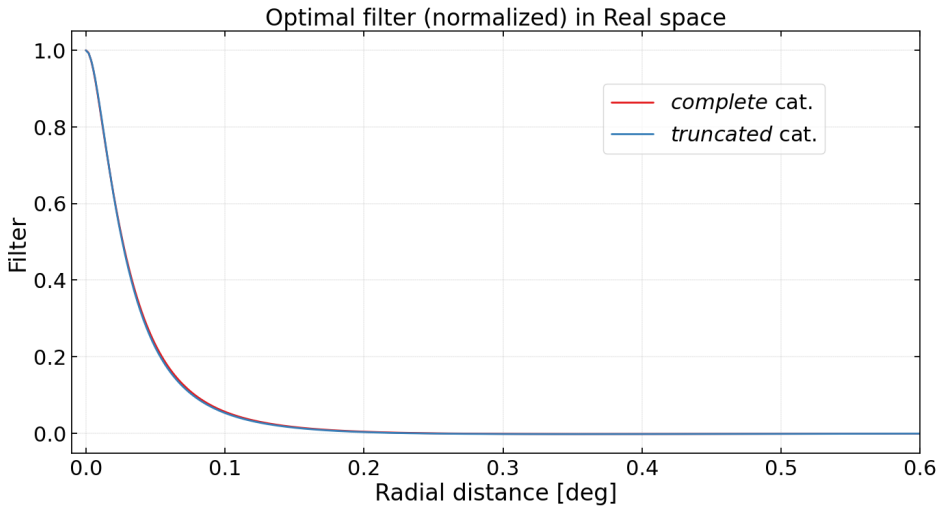


Figure 5.3 *Radial profile of the used filters in real space. They are built using the parameters reported in Table 5.2 and normalized to their maximum value. In blue we have the filter for the complete catalog, while in red the one for the truncated catalog: it is possible to notice that the two filters are very similar.*

panel in Fig. 5.5). Such ratio provides an estimate of the *expected sample purity* against pure measurement of noise as a function of SNR. The sample purity, based on this criterion, of 80% is obtained with a SNR threshold of $SNR = 1.60$.
Pixels $\boldsymbol{E-\hat{E}}$ method: the previous method is sensitive to the measurement noise fluctuations but it ignores the lensing effect due to the LSS. Looking again at the top panel of Figure 5.5 it is possible to notice an excess in the negative values distribution of the E-mode map with respect to those in the B-mode map. This excess is due to the lensing effect of the LSS where underdensities produce negative lensing peaks and we expect the same absolute value distribution for the positive peaks, which pollute the catalogs because linear density fluctuations can be well represented by a Gaussian random field as discussed in Section 2.1.3. We compare the absolute value of the negative part of the distribution of the E-mode map (see top panel of Figure 5.6) with the distribution of the positive values of the same map and compute the ratio between the two (see bottom panel of Figure 5.6). We obtain again the estimate of the *expected purity* but in this case against both measurement noise and LSS noise. Considering the same expected purity of 80%, we obtain a threshold of $SNR = 2.0$. The third and fourth methods are a variation of the previous two, where the statistics of all pixels in the map is replaced by the statistics of maxima, i.e. positive and negative peaks, see Figure 5.7. Differently from the pixel distribution, the distribution of the maxima (in the top panel of Figure 5.7) is not symmetric with respect to the zero but the E-mode map presents again a tail extending toward high values with respect to the maxima in the B-mode map. This statistics should be better suited for cluster detection because also in that case we are dealing with peaks rather than in the properties of all pixels in the map.
Peaks $\boldsymbol{E-B}$ method: the peaks, i.e. a maximum-SNR pixel, of the E- and B-mode maps are identified by AMICO through the cleaning process down to very small signal-to-noise ratios ($SNR = 1.0$); then we evaluate the values frequency

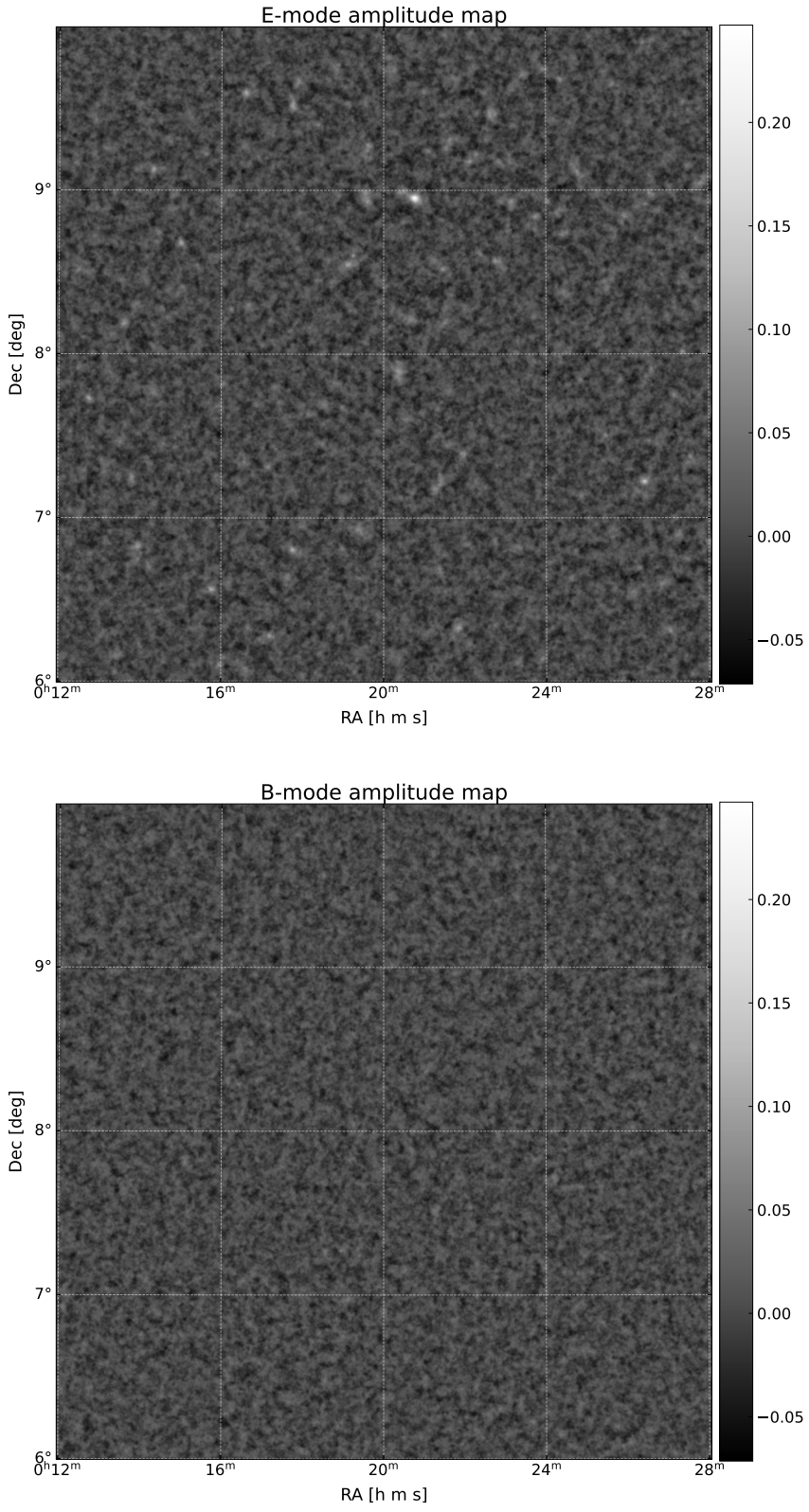


Figure 5.4 *Zoom in of the output maps in the region RA∈[3.0,7.0] deg and Dec∈[6.0,10.0]: the E-mode and the B-mode are shown in the top and bottom panels, respectively. They are plotted with the same color scale. In the E-mode map, it is possible to see the typical signatures of the presence of halos: white high-value spots. These are not present in the B-mode map, that only contains the noise from galaxies.*

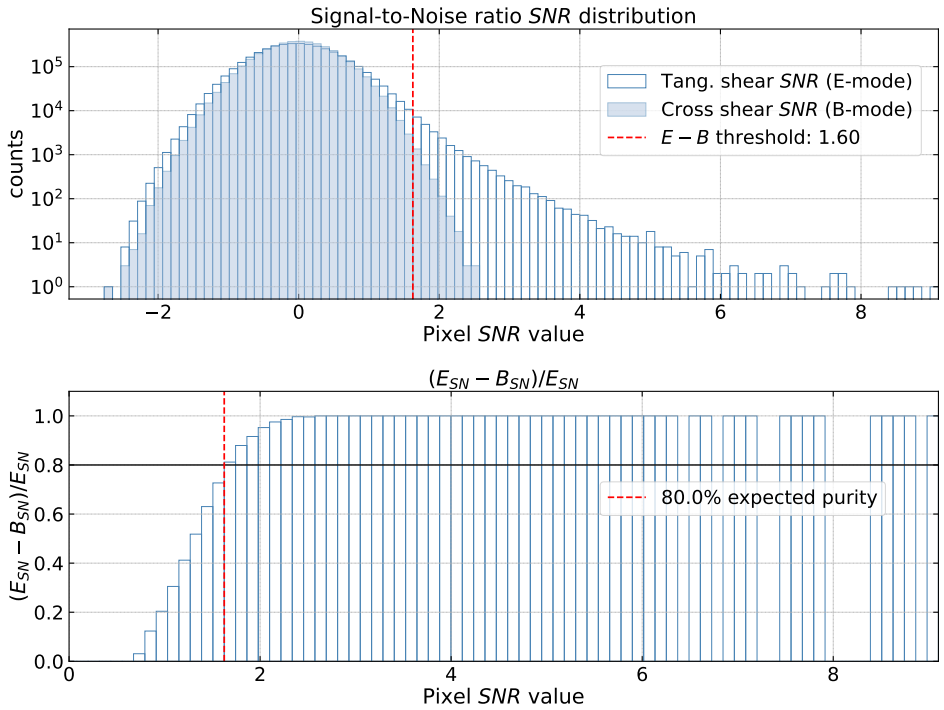


Figure 5.5 *Pixels $E-B$ method: comparison of the SNR distribution of the pixels of the E-mode and B-mode amplitude maps, and difference of the histograms relative to the truncated catalog. In the* ***top panel*** *we show the SNR distribution of the pixels of the E-mode map (empty histogram) and the B-mode map (full histogram): the B-mode map, caused only by noise, is symmetrical around 0, while the E-mode map distribution has the same symmetrical component and an additional high-value tail attributed to the lensing signal. In the* ***bottom panel*** *we have the difference of the counts in each bin of the E-mode map minus the counts of the same bin of the B-mode map, normalized over the E-mode counts. We define the threshold (red vertical line) with a given expected purity, 80%, as the SNR value in which the E-mode map counts become the 80% of the total (horizontal line); we plot the threshold also in the top Figure.*

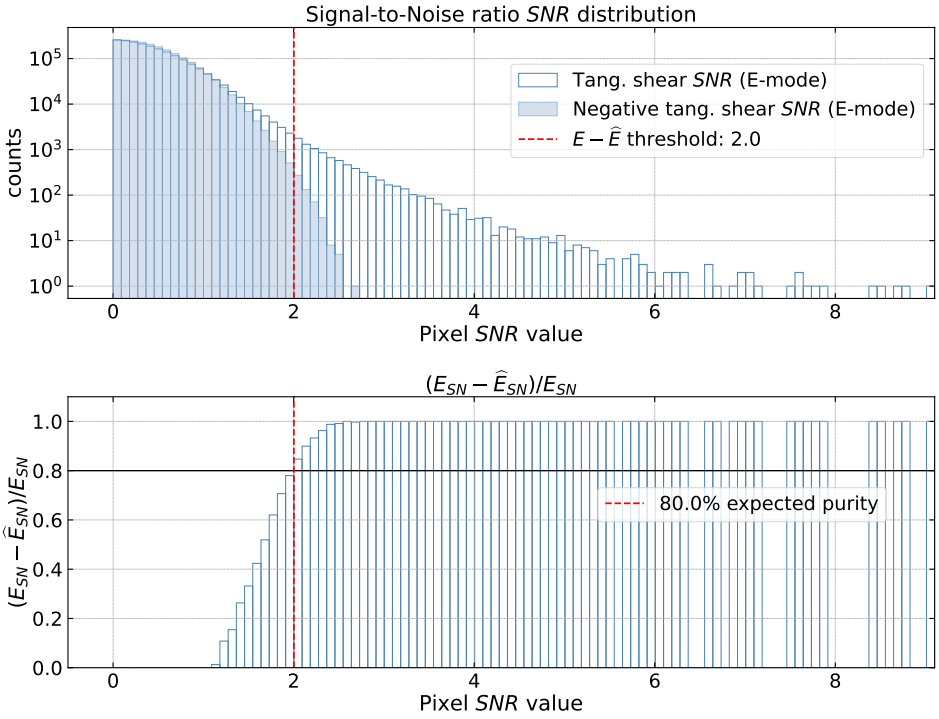


Figure 5.6 *Pixels $E - \hat{E}$ method: comparison of the SNR distribution of the positive pixels and the negative pixels of the E-mode amplitude map ($\hat{E}$), and difference of the histograms relative to the truncated catalog. In the* ***top panel*** *we show the SNR distribution of the positive pixels of the E-mode map (empty histogram) and the negative pixels after having changed the sign, $\hat{E}$ (full histogram), the negative pixels are caused by the noise and by the negative signal of the lensing effect that is not related to galaxy clusters. In the* ***bottom panel*** *we have the difference of the counts in each bin of the positive and negative pixels of the E-mode. We define the threshold (red vertical line) with a given expected purity, 80%, as the SNR value in which the positive pixels map counts become the 80% of the total (horizontal line); we plot the threshold also in the top Figure.*

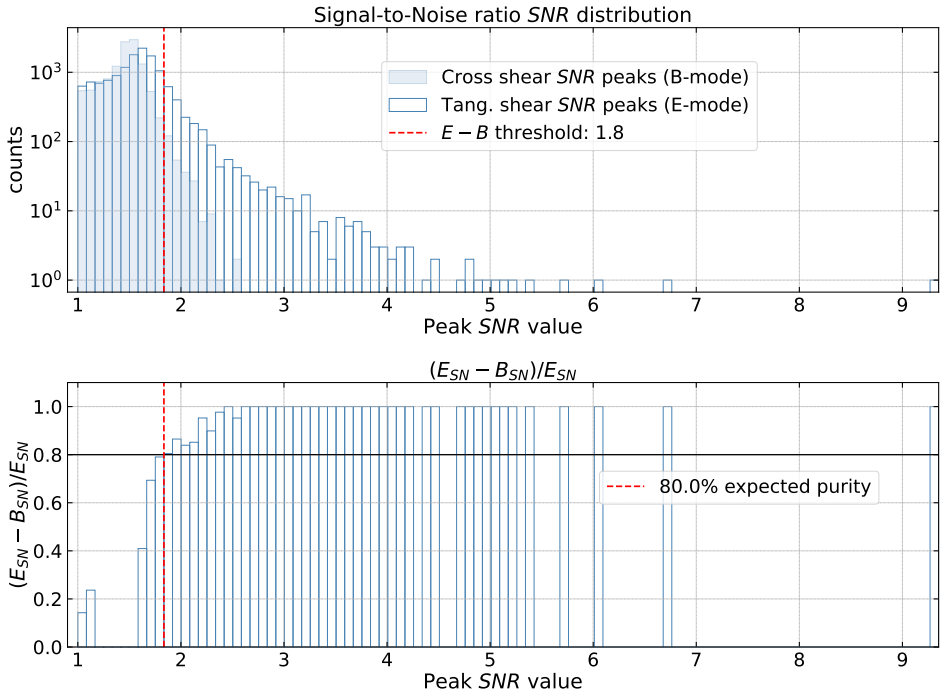


Figure 5.7 *Peaks $E-B$ method: comparison of the SNR distribution of the peaks of the E-mode and B-mode amplitude maps, and difference of the histograms relative to the truncated catalog. In the* ***top panel*** *we show the SNR distribution of the peaks of the E-mode map (empty histogram) and the B-mode map (full histogram). The peaks of the B-mode map, caused by noise only, reach SNR values that are quite lower with respect to the E-mode peaks. In the* ***bottom panel*** *we have the difference of the counts in each bin of the E-mode map minus the counts of the same bin of the B-mode map, normalized over the E-mode counts. We define the threshold (red vertical line) with a given expected purity, 80%, as the SNR value in which the E-mode map counts become the 80% of the total (horizontal line); we plot the threshold also in the top figure as a vertical red line.*

and distributions in full analogy to what has been done for the pixel statistics (see bottom Figure 5.7). The expected purity of 80% is found for a threshold of $SNR = 1.8$. **Peaks $\boldsymbol{E-\hat{E}}$ method**: also the last method presented does not account for the LSS noise contribution, therefore, as done for the **Pixels $\boldsymbol{E-\hat{E}}$ method**, we now consider the distribution of the negative peaks. Again, we perform a detection run for the E-mode map and the E-mode map with inverted sign (so that now the minima are maxima) with a threshold of $SNR = 1.0$. We compare the distributions of maxima and minima (see top panel of Figure 5.8). Then we compute the ratio of the distribution obtaining again an estimate of the expected purity, for the peaks themselves. For 80% of expected purity, we derive a threshold of $SNR = 2.4$. Since the peak statistics is better suited for cluster detections as discussed, we will consider the methods based on it in the following analysis. Table 5.3 lists the SNR thresholds derived with both **Peaks $\boldsymbol{E-B}$ method** and **Peaks $\boldsymbol{E-\hat{E}}$ method** methods for the two catalogs (full and truncated). We will refer to the threshold found with the peaks $E-B$ method as the *weak threshold* while the

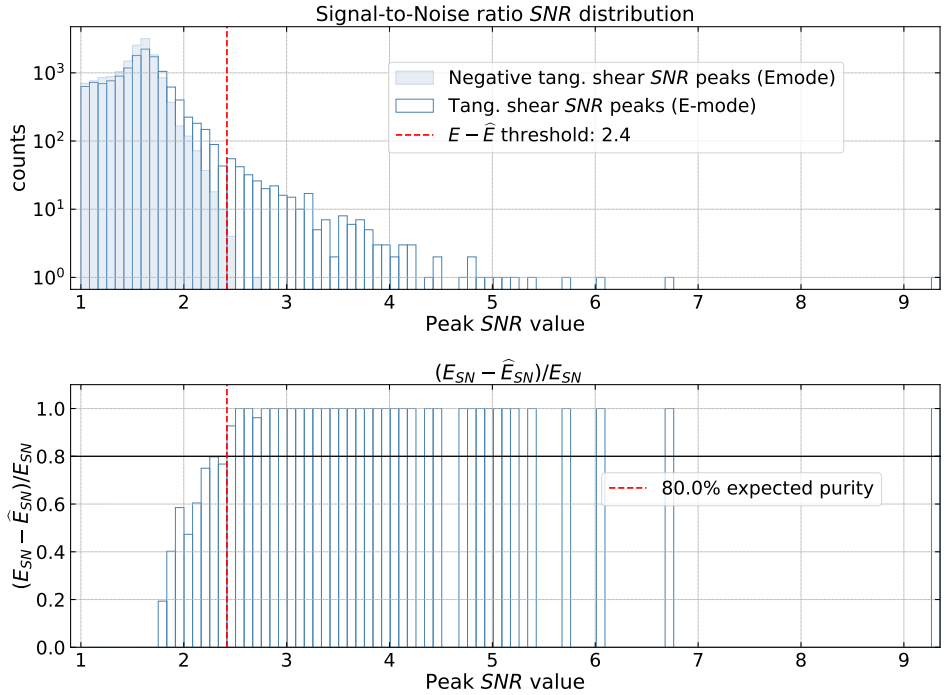


Figure 5.8 *Peaks $E - \hat{E}$ method: comparison of the SNR distribution of the positive peaks and the negative peaks of the E-mode amplitude map ($\hat{E}$), and difference of the histograms, relative to the truncated catalog. In* ***top panel*** *we show the SNR distribution of the positive peaks of the E-mode map (empty histogram) and the negative peaks in absolute value, $\hat{E}$ (full histogram). The negative peaks are caused by the noise and by the negative signal of the lensing effect that is not related to galaxy clusters. In the* ***bottom panel*** *we have the difference of the counts in each bin of the positive and negative peaks of the E-mode. We define the threshold (red vertical line) with a given expected purity, 80%, as the SNR value in which the positive pixels map counts become the 80% of the total (horizontal line); we plot the threshold also in the top figure as a vertical red line.*

catalog	weak SN_{thr} $E-B$ method	strong SN_{thr} $E-\hat{E}$ method
complete	1.9	**2.4**
truncated	1.8	**2.4**

Table 5.3 *Detection thresholds used for the analysis of the complete and truncated catalog of simulated ellipticities. For each catalog, we report the thresholds computed with the peaks analysis of the E-mode and B-mode peaks (peaks $E-B$ method) and the peaks analysis of the positive and negative peaks of the E-mode (peaks $E-\hat{E}$ method).*

$E-\hat{E}$ method produces the *strong threshold*.

5.3.2 Catalog of detections

We identified with AMICO all peaks above the SNR thresholds found in Section 5.3.1 and produced the results for both full and truncated galaxy samples. For the truncated galaxy catalog, we have 1796/329 objects for the weak/strong SNR thresholds. Roughly 18% of the detections found with the weak threshold is above $SNR = 2.4$. For the complete galaxy catalog, we have 1189/268 objects for the weak/strong SNR thresholds. As expected we have more detections based on the truncated galaxy catalog which comprises fewer foreground objects reducing the lensing signal.

In Figure 5.9 we show a zoom in of the amplitude map (E-mode) where we marked the real detections. The peaks that are more visible in the map are have a higher SNR as expected.

In the top panel of Figure 5.10 we show the SNR of the detections as a function of their ID which marks the order in which they are discovered. The same figure shows the two SNR thresholds adopted in the analysis. In the bottom panel of Figure 5.10 we display the signal-to-noise ratio distribution of the detections, the median SNR of the detection catalog found with the weak threshold is 2.04 ± 0.14 [2] while the median of the sample of detections obtained with the strong threshold is 2.8 ± 0.3. Clearly, the majority of the detections have a low signal-to-noise ratio building $\sim 78\%$ of the total.

5.4 Matching procedure and weak lensing selection function

When dealing with detection algorithms or in general with the creation of a catalog of astrophysical objects it is useful to compare the results against an existing catalogs of the same type of sources present in the same FOV to check for consistency. This can also allow to extend the information about the detected sources, for instance by exploiting different wavelengths and methods. In the case of simulated data, it is available a reference catalog listing the actual sources in the simulation with all

[2] From here on we will express the error associated with the median using the *median absolute deviation* (MAD), a robust measure of the variability of a univariate sample of quantitative data: $MAD = \text{median}(|X_i - \tilde{X}|)$ where $\tilde{X}$ is the median of the variable X.

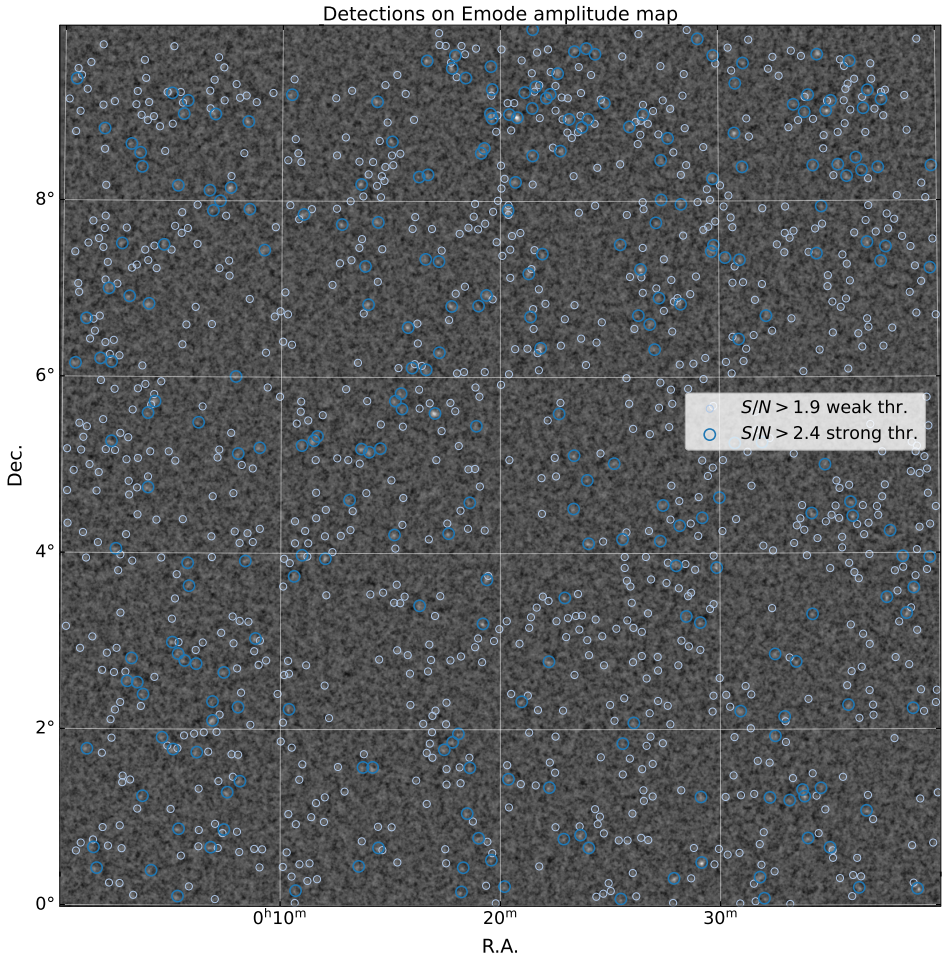


Figure 5.9 *Map of the detections from the truncated catalog. We show the positions of the detections on the amplitude map where. We use the bigger blue circles for the detections with $SNR > 2.4$, the strong threshold, while the smaller light-blue circles correspond to the detections with $SNR > 1.8$, the weak threshold.*

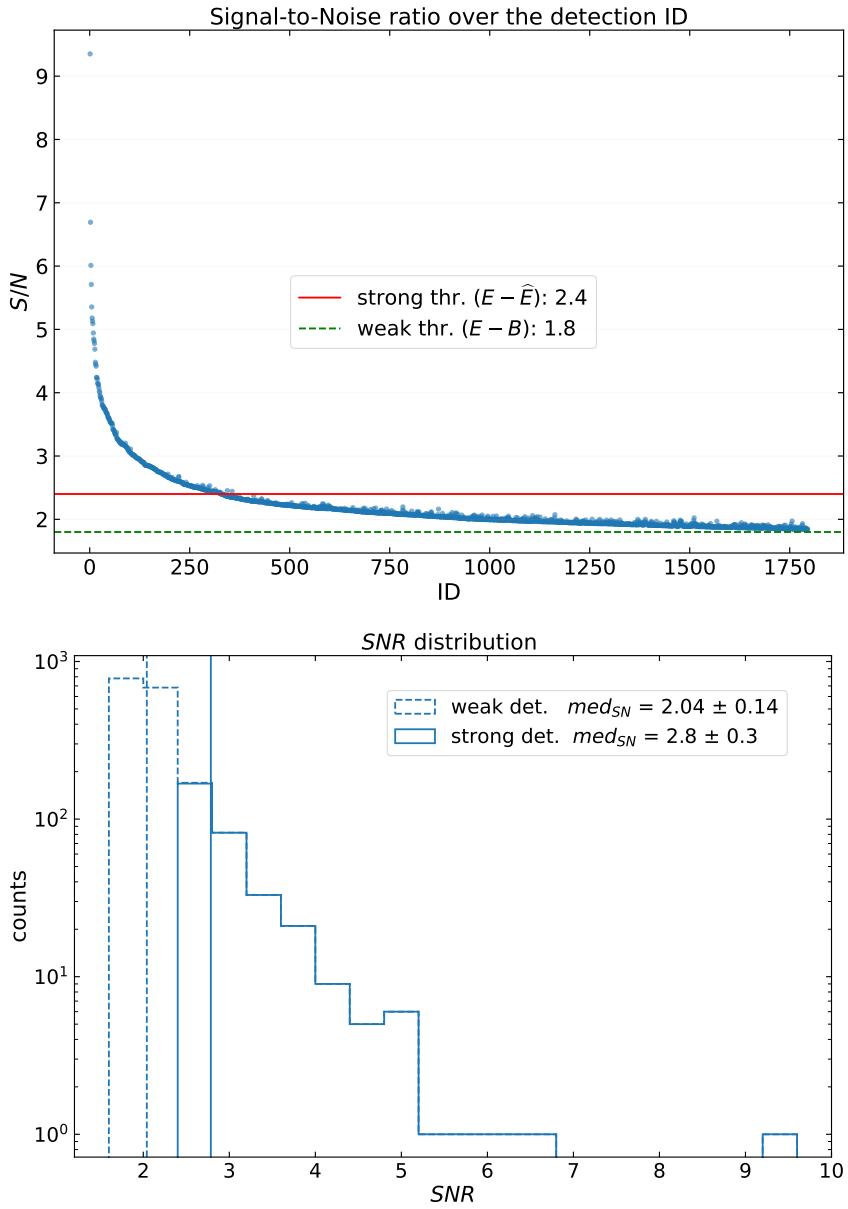


Figure 5.10 *Signal-to-noise ratio analysis of the catalog of detections. In* ***top panel*** *we show the SNR of the detections as a function of their ID in the catalog, which can be used to follow the different steps of the detection procedure which starts from the higher SNR that is ∼9.0 and then rapidly decreases reaching $SNR \sim 3.0$ around ID∼100. Below the strong threshold (solid red horizontal line) the rare jumps due to the cleaning become more common. The green dashed line represents the weak threshold. In* ***bottom panel*** *we show the histogram of the SNR distribution of the detection catalog, the counts are in logarithmic scale; with the solid line we enhance the distribution of the peaks with $SNR > 2.4$, the subset found with the strong threshold, while the dashed line is the full sample of detections. The vertical lines stand for the median of the distributions.*

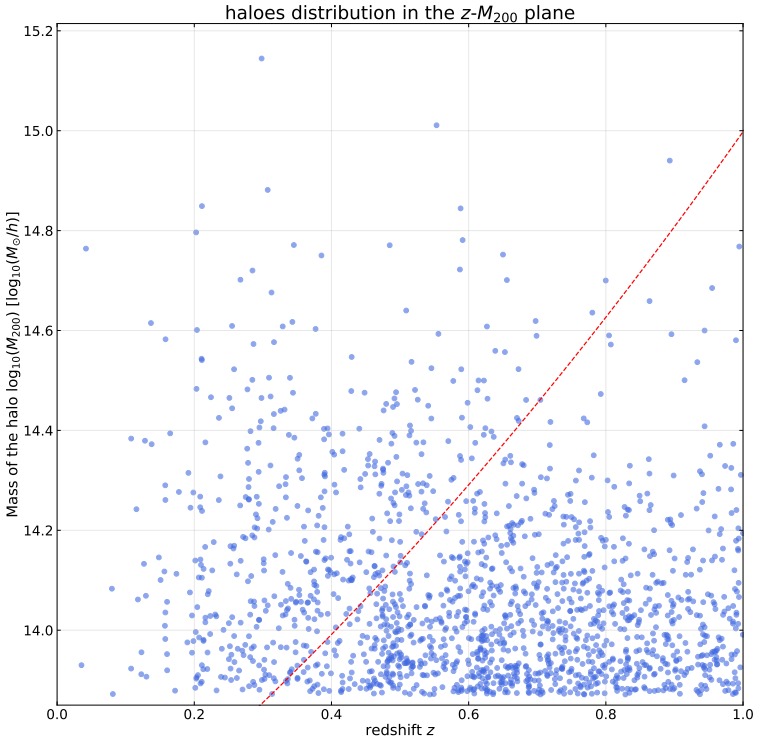


Figure 5.11 *We show the distribution of the complete catalog of dark matter simulated structures in the redshift-mass plane. The red dashed line reproduces the cut applied, only the halos above the line are used as a reference as they are expected to produce a SNR high enough to be detected. Also, possible matches of detections with halos in the region under the red line would likely result from a random superposition in the parameter space $z - M$ because of the high density of sources.*

their true properties. This catalog allows to test the reliability of the code and in general of the overall procedure. This means that for weak lensing we know where the structures generating the lensing peaks are. We will refer to this catalog as the *halo catalog* to be matched with the detections catalog to characterize the sample purity and completeness.

For the matching, we define a subset of true halos to reduce the confusion given by random projections that weak lensing measures can not address being unable to provide an estimate of the redshift. We first select all halos within the FOV. Then we apply a mass dependent selection which scales with redshift to exclude all clusters that cannot produce a significant lensing signal and that could be matched only because of a chance projection. Such cut is defined by modeling the expected signal-to-noise ratio of the sources assuming a survey with a certain surface number density. We adopt the same selection criterion used within the weak lensing detection challenge currently in place within the Euclid consortium. The cut is shown in Figure 5.11. Only structures above this line are considered. This criterion is meant to reduce the ambiguity given by chance projections.

The final reference catalog used for matching comprises 464 dark-matter objects. In Figure 5.12 we show the map of these halos, where the color of the points indicates

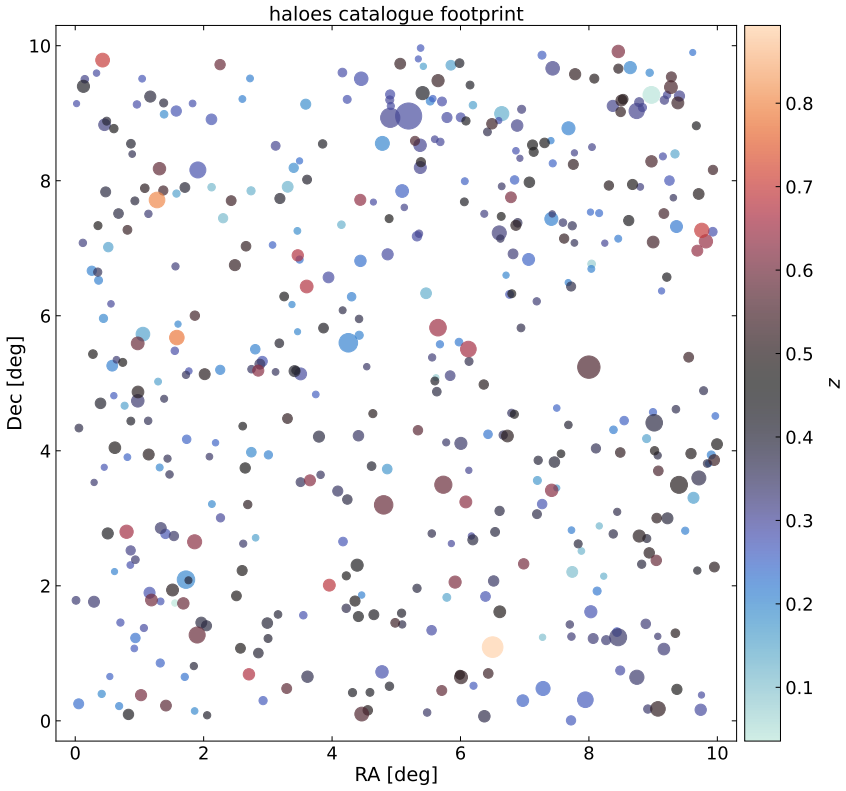


Figure 5.12 *Map of the dark matter halos used as a reference for the detection. We show the halos selected from the simulated halo-catalog plotted in the FOV. The color of the dots is related to the redshift and the size is scaled with the virial mass. It is possible to see some cases of aggregated structures where we have circles of similar colors near each other.*

the redshift and their size scales with the mass. The distributions of redshift and mass of all halos and of the selected ones are shown in Figure 5.13. Most of the halos are in the redshift range [0.2,0.5] with a median of 0.35 ± 0.09, and a virial mass with a median of 14.2 ± 0.16 $\log(h^{-1}\mathrm{Mpc})$. The adopted selection removes most of the sources with redshift above $0.5 - 0.6$ that would be impossible to detect because of their high redshift and small mass.

We performed the matching analysis using a code available within the AMICO software package. The code sorts the lensing detections according to their signal-to-noise ratio and the halos with respect to their virial mass. This is done to match detections with large SNR to large clusters. The *matching radius* is set to 1 h^{-1}Mpc centered on the true position of the halo, and its angular size scales according to the angular diameter distance of halos. This choice is based on the typical size expected for the weak lensing imprint in the data and is large enough to account for the discretized position of halos in the simulations and detections in the amplitude maps. A larger matching radius would create an artificial increase of successful matching due to chance alignments. We also set a higher limit of 3 arcmins on the matching radius in order to stop the increase of the radius for sources at high redshift.

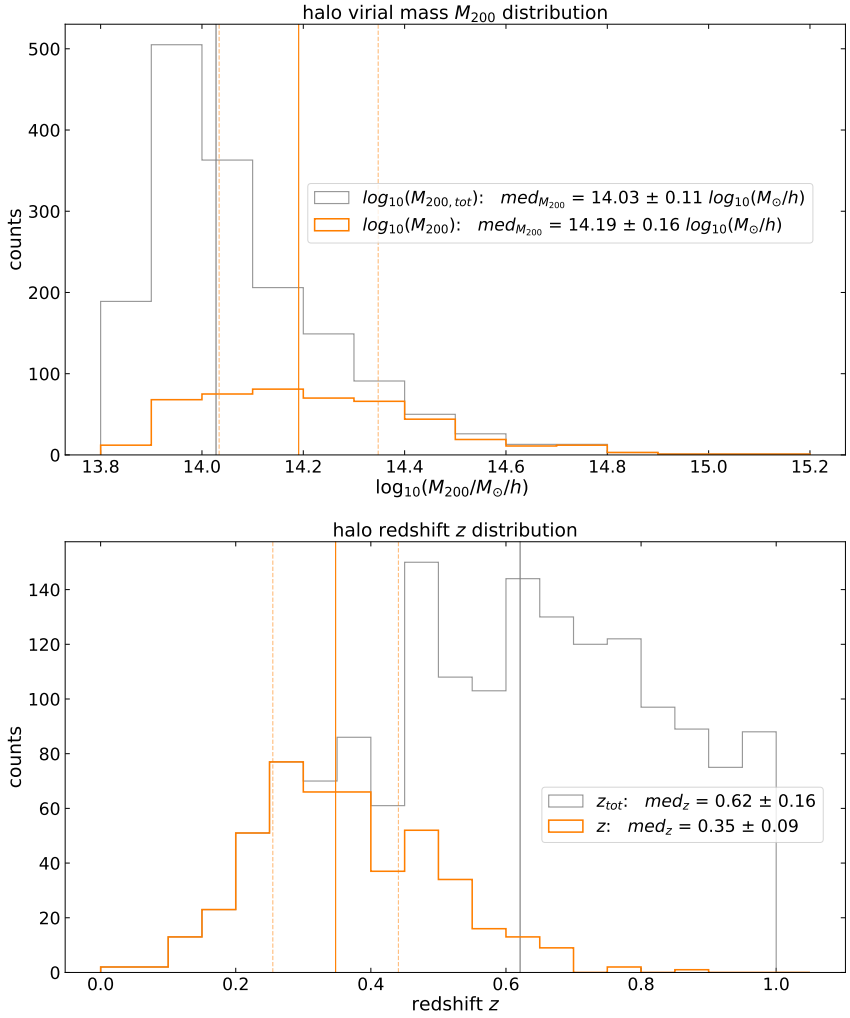


Figure 5.13 *Distribution of virial mass (M_{200}) and redshift of the 464 selected sources from the simulated dark matter halos in the "Euclid test challenge" halo catalog. In **top panel** we show the distribution of virial mass in $\log_{10}(M_{200})$ of the selected halos (in orange) compared with the distribution of the total halos from the simulated catalog (in gray), the selected catalog has a median of 14.19 ± 0.16 $\log_{10}(h^{-1}M_\odot)$ (error in median absolute deviation) while the total catalog has a lower median of 14.03 ± 0.11 $\log_{10}(h^{-1}M_\odot)$. In **bottom panel** we compare the redshift distributions between the total and the selected catalog.*

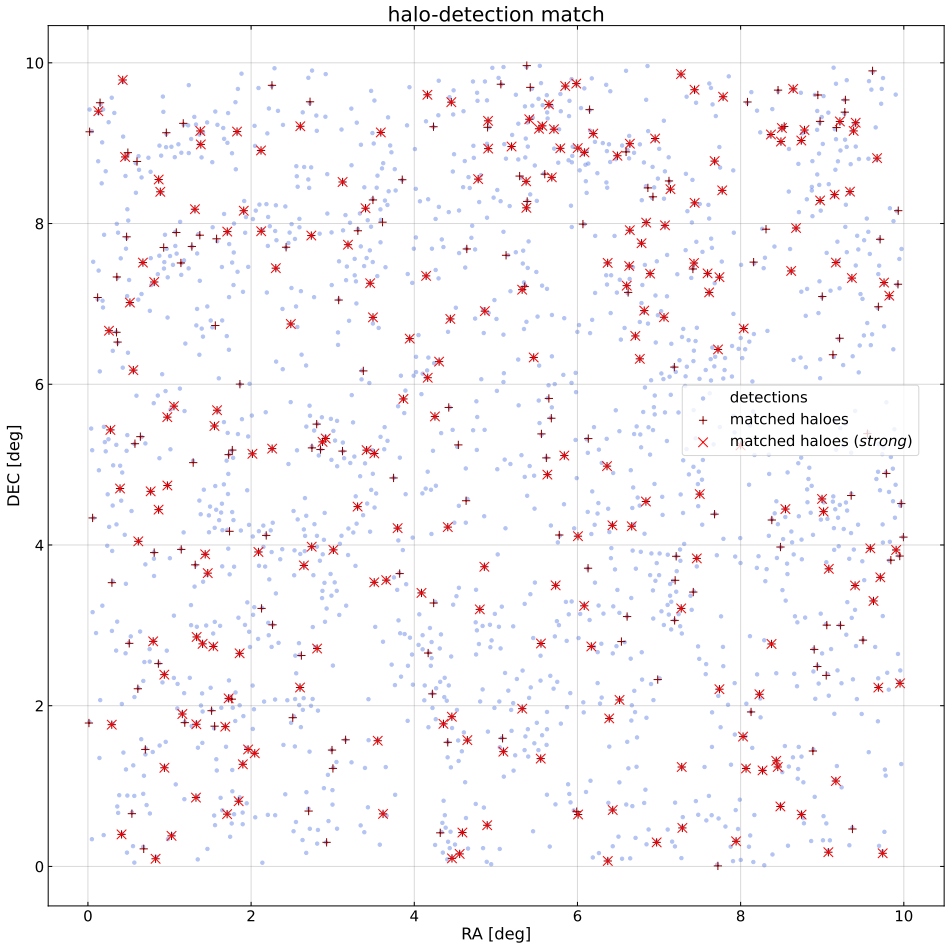


Figure 5.14 *Distribution of the matched detections. We show in light blue all the detections of the complete catalog. The red plus symbol stands for the matched halo with detection in the weak threshold catalog ($SNR > 1.8$) while the bigger red cross symbol is associated only with the detections in the strong threshold catalog ($SNR > 2.4$).*

5.4.1 Matching analysis

The catalog of the matched detections for the larger sample, obtained with the weak threshold, contains 323 objects, while the strong threshold sample of detection match 187 dark-matter halos of the 464 in the catalog described in Section 5.4. The distribution of the detected and matched clusters can be seen in Figure 5.14.

Figure 5.15 shows the distribution of the separation between detections and the matched clusters in h^{-1}Mpc (top panel) and arcminutes (bottom panel). The figure refers to the truncated galaxy catalog for both the weak ($SNR = 1.8$) and strong ($SNR = 2.4$) thresholds.

In terms of angles, the median displacement of r of the full catalog of detections is of 0.5 ± 0.3 arcmin, see bottom panel of Figure 5.15. For the sample of matched detections with $SNR > 2.4$, the median is lower as these matches are less likely to be random, we have 0.4 ± 0.2 arcmin. The fact that the detections-matches

displacement is close to the pixel shows once more the validity of the matching procedure. Only the small tail with larger displacements could be attributed to random matches.

The displacement distribution in h^{-1}Mpc (see top panel of Figure 5.15) shows that the displacement is indeed limited to small separations with a median displacement of $\sim 0.1h^{-1}$Mpc with respect to the matching radius of 1 h^{-1}Mpc. Exactly we have a median displacement of $0.13 \pm 0.07h^{-1}$Mpc for the full sample ($SNR > 1.8$) and of $0.11 \pm 0.05h^{-1}$Mpc the subset with $SNR > 2.4$.

Figure 5.16 shows the distribution of matched halos as a function of mass and redshift. We use a color scale to represent the signal-to-noise ratio of the matched detections. For the sake of clarity, the matched halos from the $SNR = 2.4$ threshold subset are represented with a cross symbol while the full sample is represented with dots. The gray dots on the other hand stands for the halos of the total simulated catalog.
Most of the detections with higher SNR are found in the redshift range $[0.2, 0.4]$ and the logarithm of the corresponding halos with mass over 14.4 $M_\odot/h$. Most of the un-matched halos are found in regions with low mass and intermediate redshift or higher mass but at higher redshifts, i.e. the regions where the SNR is the smallest. Therefore, only the detections found with the lowest SNR threshold will be present in these regions. Some care has to be taken when dealing with these low SNR detections but as discussed the matching criteria we implemented seem to be solid.

5.4.2 Signal-to-noise ratio

The distribution of the SNR is shown in Figure 5.17, with a median of 2.5 $\pm$ 0.4 and 2.9 $\pm$ 0.3 for the catalogs of matched detections found with $SNR > 1.8$ and $SNR > 2.4$ respectively. The dashed line shows the distribution of matched detections, the solid line shows the distribution of the high-SNR subset, and the solid shaded background the distribution for all detections. With the vertical lines we show the medians of the distributions, dashed for the full sample, and solid for the high-SNR. More than 90% of the *un-matched* detections of the weak catalog have $SNR < 2.4$. This shows that the threshold obtained from the $E - B$ method with an expected purity of 80% is not reliable.

5.4.3 Redshift and virial mass

The two main properties related to weak lensing detections are the (virial) mass of the halo and its redshift. We show in Figure 5.18 the distribution with respect to mass and redshift of all selected halos (shaded region), the matched halos with the sample of detections (dashed histogram) and the subset of matched halos related to a high-SNR detection (solid histogram). We show the medians of the matched halo distributions with vertical lines of the same style, dashed for $SNR > 1.8$ and solid for $SNR > 2.4$.

The virial mass of the matched halos is distributed with a median of 14.23 ± 0.16 $\log_{10}(M_\odot)$, and of 14.29 ± 0.15 $\log_{10}(M_\odot/h)$ for the high-SNR subset. Not every mass can be detected with the same efficiency, clearly, the fraction of halos with larger masses that are matched is higher but not every massive halo is detected, as we can see in the high mass tail. For example, the third most massive halo appears not to be associated with a lensing detection (see Section 5.4.4): in this case, the

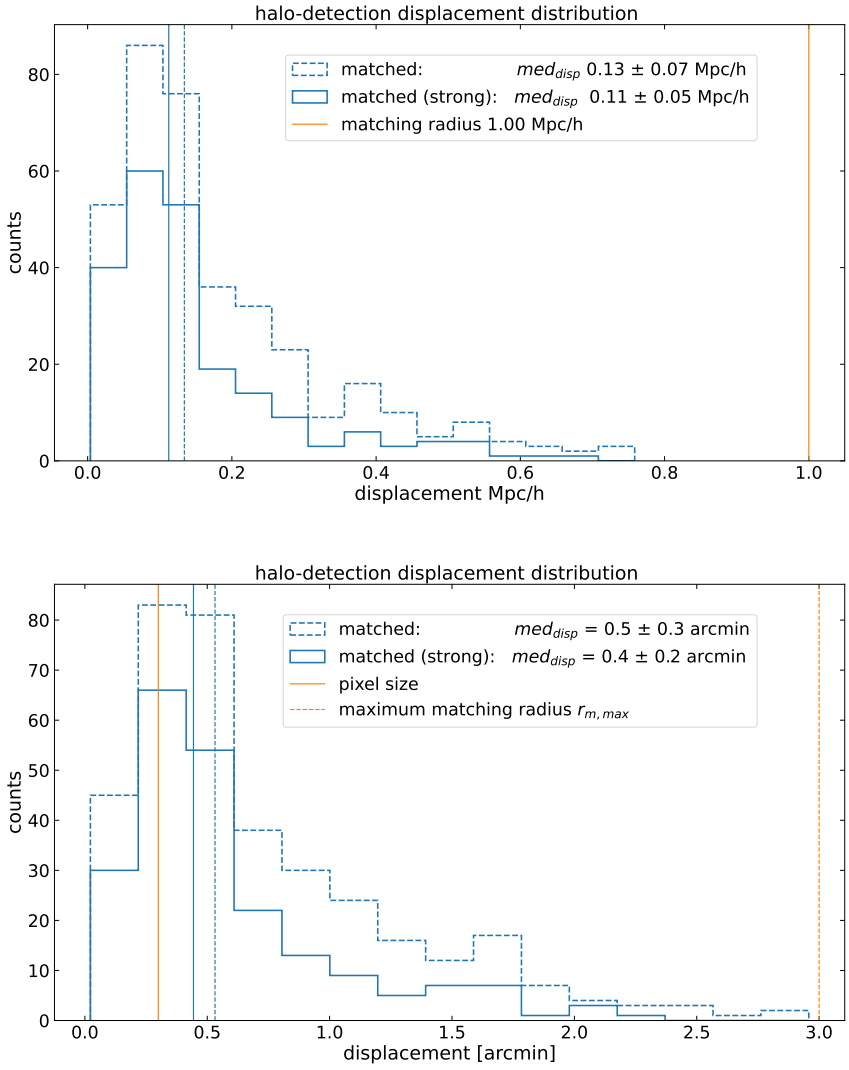


Figure 5.15 *Halo-Detection displacement distribution. In the* ***top panel*** *we show the distribution of the spatial distance between the position of the halo and the position of the detection that matches the halo. We expect the position of the halo and the corresponding detection to be slightly different due to the use of grids in AMICO-WL. We show the medians of the distributions (vertical lines) and the 1 h^{-1}Mpc matching radius (orange vertical line). In the* ***bottom panel*** *we show the distribution of the displacement in the projected plane. The vertical orange line shows the dimension of the pixel used in AMICO-WL map creation. In both figures the dashed line represents the matched detections using the weak threshold while the solid represents the subset of matched detections using the strong threshold.*

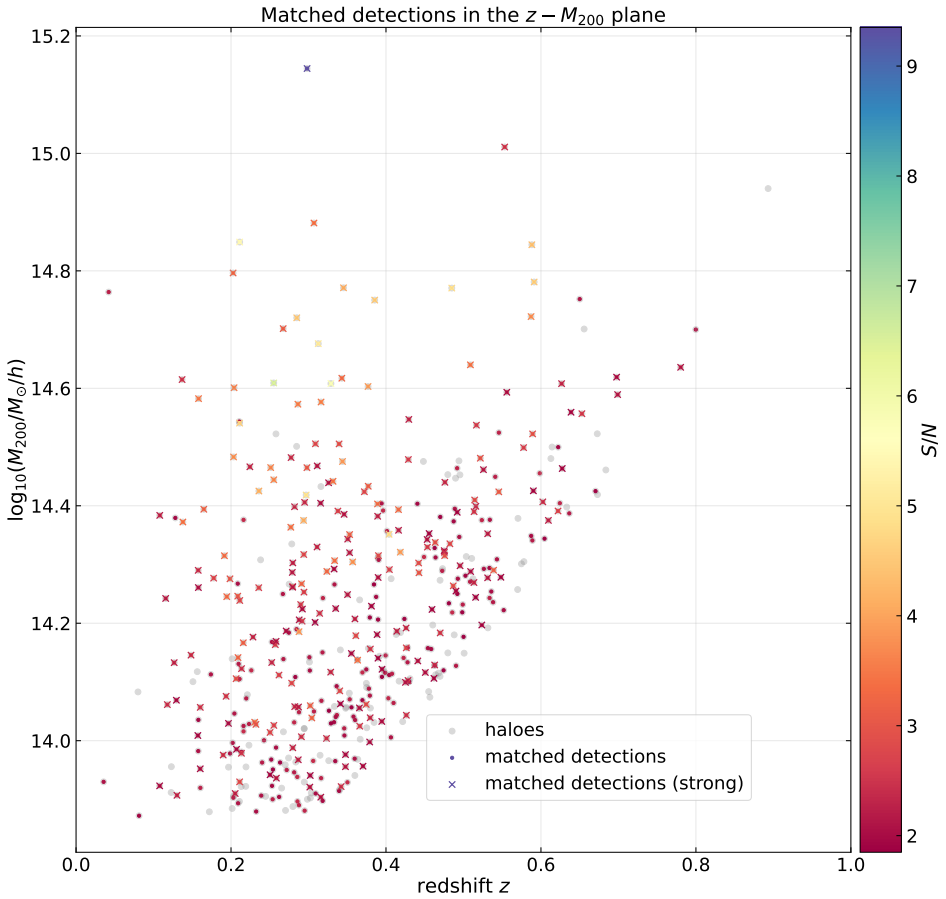


Figure 5.16 *Distribution of the halos as a function of Mass and Redshift. Grey dots represent all the 464 simulated halos. The colored dots are the full sample of matched halos, while the ones in the $SNR > 2.4$ subset are highlighted with a cross symbol. The color scale of the dots and cross symbols represents the SNR of the detections. The halos in the regions with high mass, $\log_{10} M_{200}/M_\odot)/h > 14.3$, and redshift $\approx 0.2 - 0.4$ are mostly matched and with high-SNR detections.*

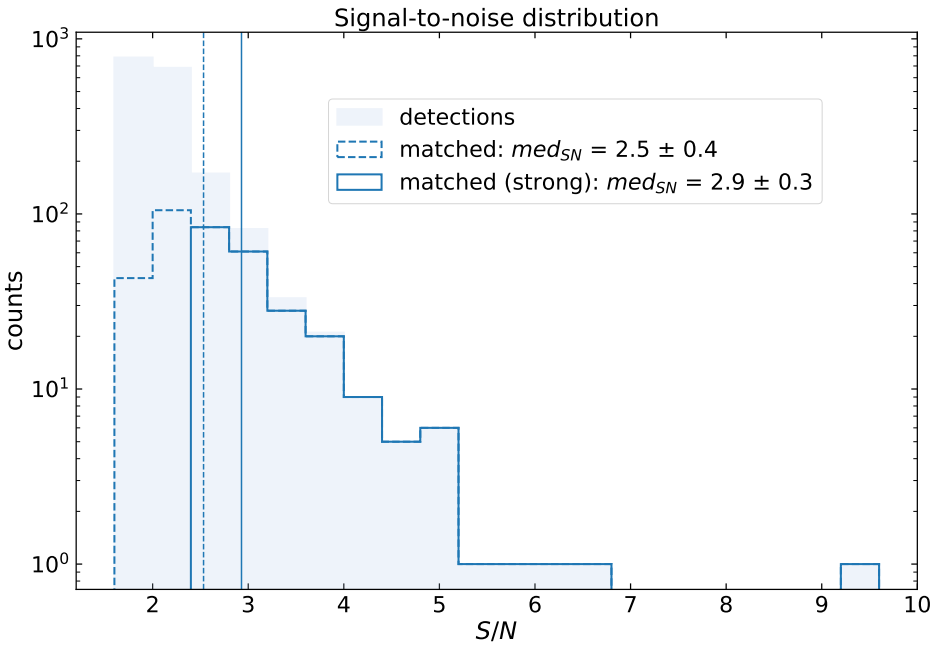


Figure 5.17 *Signal-to-noise ratio distribution of the matched detections. We show the distribution of the SNR for the weak threshold detections (dashed histogram) and the subset of the strong threshold detections (solid histogram); we show the medians with vertical lines of the same style. The background shaded histogram represents the SNR distribution of the whole detection catalog. It is clear that most of the un-matched detections have a low signal-to-noise ratio.*

detection associated exists but the position of its maximumis slightly outside the adopted 3 arcmin matching radius.

For the redshift, the medians of the matched halo distribution are 0.34 ± 0.09 (see Fig. 5.11) and very similar for the high-SNR subset. The distribution conveys a different efficiency in the detection of clusters at different redshift. As expected, the filter selects preferentially the halos that have an intermediate distance between the observer and nearby background sources, i.e. where the lensing signal is the strongest.

5.4.4 The un-matched halos

It is crucial to study the tendency of the detection algorithm to produce false negatives. Therefore we select a sample of un-matched halos so that we can focus on the most striking cases. We show the sample of un-matched halos studied in Figure 5.19. Other objects with similar mass and redshift have been detected with high significance, therefore we would expect to match these objects as well but this is not the case.

Some missed matched halos are easily explained. We refer to the Figure 5.20 where we show the amplitude map with the detections (blue and light-blue circles), the matched halos (red dots), and the un-matched selected halos (yellow stars with labels).

- The object labeled with **1** (see Figure 5.19) has a large mass of $14.7\ \log_{10}(\mathrm{M}_{\odot}/h)$ but itis at relatively high redshift $z = 0.66$, other detections associated to similar objects have a typical SNR below 2.4.
- In the case of halo **6**, it is likely that the algorithm fails to map the signal of a halo at the edge of the FOV: this could lead to a low SNR detection, an uncorrect position of the peak, or no detection at all. In this case, there is a low-SNR detection nearby but it is too far.
- Halo **18** and halo **12** are large haloes near other two, smaller, halos that instead have been matched; it is possible that those other halos should have been matched with nearby lower SNR peaks but instead matched with higher SNR peaks, leaving **18** and **12** un-matched.
- Halos labeled with **3**, **7**, **11**, and **17** are placed on the map near extended regions of high amplitude, may be related to a complex dark matter structure or high noise fluctuations. In these cases, the detection algorithm found several low-SNR peaks instead of one big peak resulting in an inefficient detection of the matter overdensity then a missed matching.
- By inspecting the amplitude map, there are three cases, **5**, **8** and **9**, in which there is one (or two in the case of **8**) clear detection with high significance that lays just outside the maximum search radius of 3 arcminutes adopted during the matching. In these circumstances, the failure is not in the detection algorithm but in the matching procedure. All of these three detections are within 6 arcminutes.
- Halo **2** and halo **15** represent two peculiar cases as they have very high mass and redshift around 0.3. The lensing signal should be strong but no relevant detection is visible in the amplitude maps.

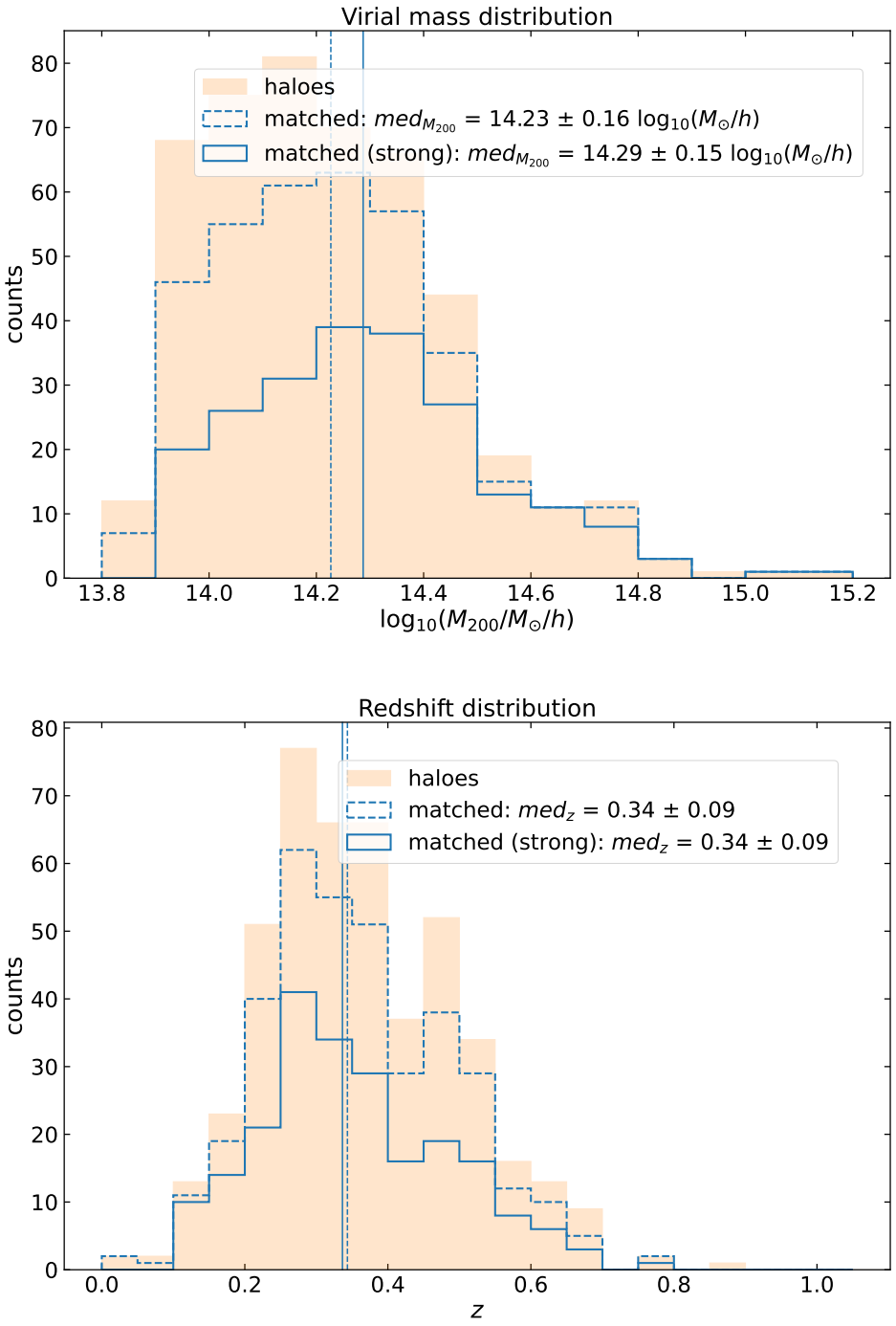


Figure 5.18 *In **top panel** we show the distribution of virial mass (in logarithm) while in **bottom panel** we have the redshift distribution for the matched halos. The background shaded orange histogram is the distribution of the total 464 simulated dark matter halos. In both panels, the dashed line refers to the full sample, and the solid line to the high-SNR subset.*

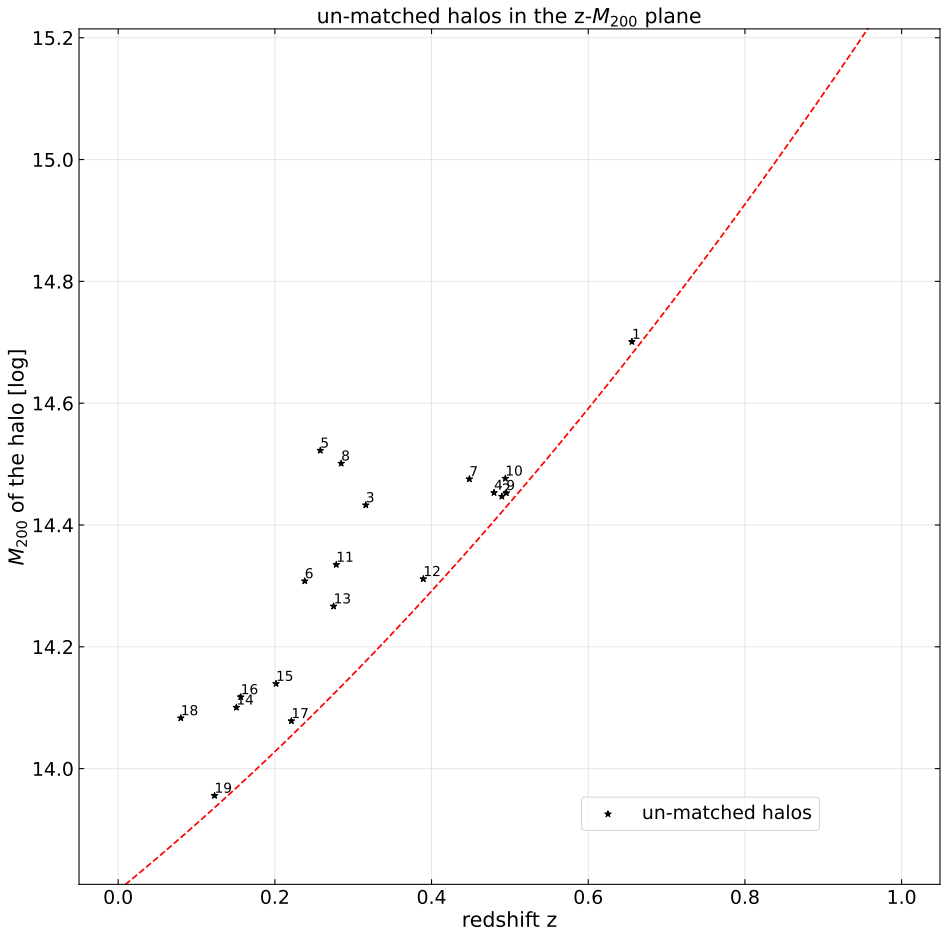


Figure 5.19 *Distribution of the un-matched halos as a function of Mass and Redshift. We used a more stringent selection function (red dashed line) to focus on the un-matched halos which are associated with a high expected SNR. Grey dots are all the 464 simulated halos. The un-matched selected halos are black stars.*

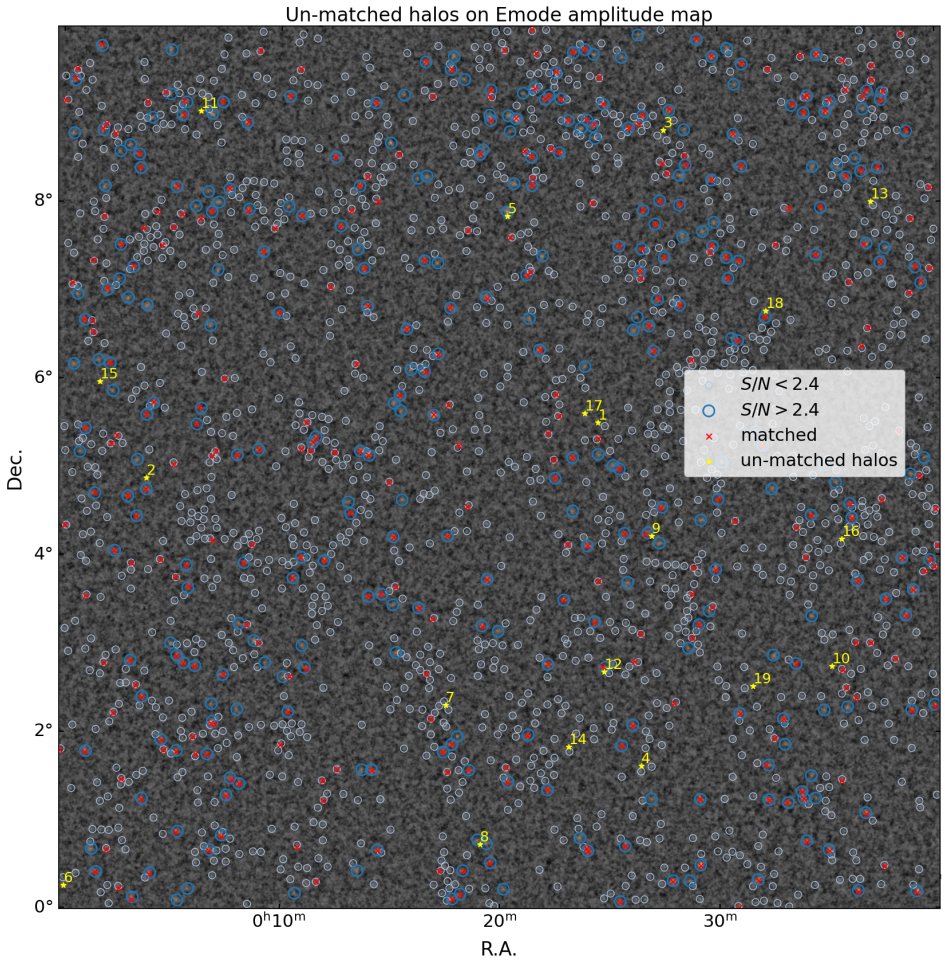


Figure 5.20 *Map of the un-matched selected halos. We show the amplitude map with the low-SNR detections as light-blue circles, high-SNR detections as blue circles, the matched halos as red cross symbols, and the un-matched halos discussed in the text are yellow stars.*

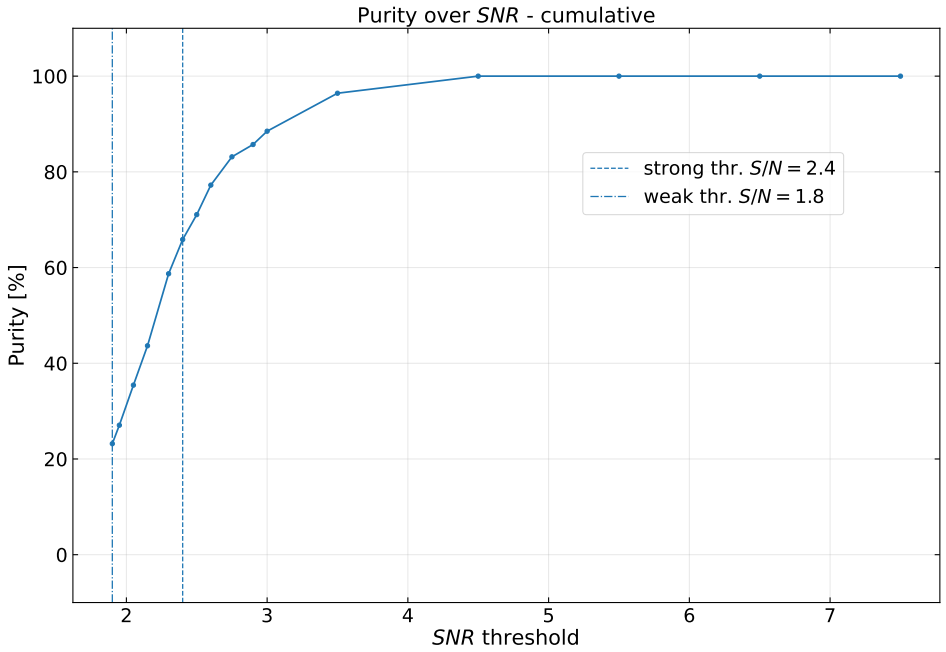


Figure 5.21 *Cumulative purity of the sample over different signal-to-noise ratio thresholds. We show the purity of the catalog of detections over different values of SNR, from $SNR = 8.0$ down to the weak threshold of $SNR = 1.8$. With the vertical lines, we highlight the two thresholds corresponding to the weak and strong criteria for the detection.*

5.5 Evaluation of the sample purity and completeness

With the matching at hands, we now evaluate the sample *purity* and *completeness*. The first is the fraction of detections that match with a real source over the total number of detections, while the latter is the fraction of matched detections over the total number of halos. We use the halos above the red curve in Figure 5.11 as a reference catalog for matching.

In Figure 5.21 we show the purity for the sample above a certain SNR. Clearly, the purity is lower for lower SNR and reaches a value of 80% very close to the strong threshold identified with the method *peaks* $E - \hat{E}$, actually 70%. This shows the validity of the method. The sample derived with the threshold of the *peaks* $E - B$ method has a very low purity of 20%, further proving that the noise contribution of the LSS can not be neglected.

In Figure A.13 we show the *completeness* of the sample of detections i.e. the fraction of matched halos over the total number of halos, as a function of SNR. The completeness decrease with the increasing of the SNR as the sample of detections decreases in number. With the relaxed threshold, $SNR = 1.8$, the completeness is around 80%, while the strong threshold, $SNR = 2.4$ determines a completeness of 40%.

The sample completeness is shown as a function of SNR in Figure 5.23 and for different SNR thresholds. As expected, by applying a very restrictive/relaxed SNR

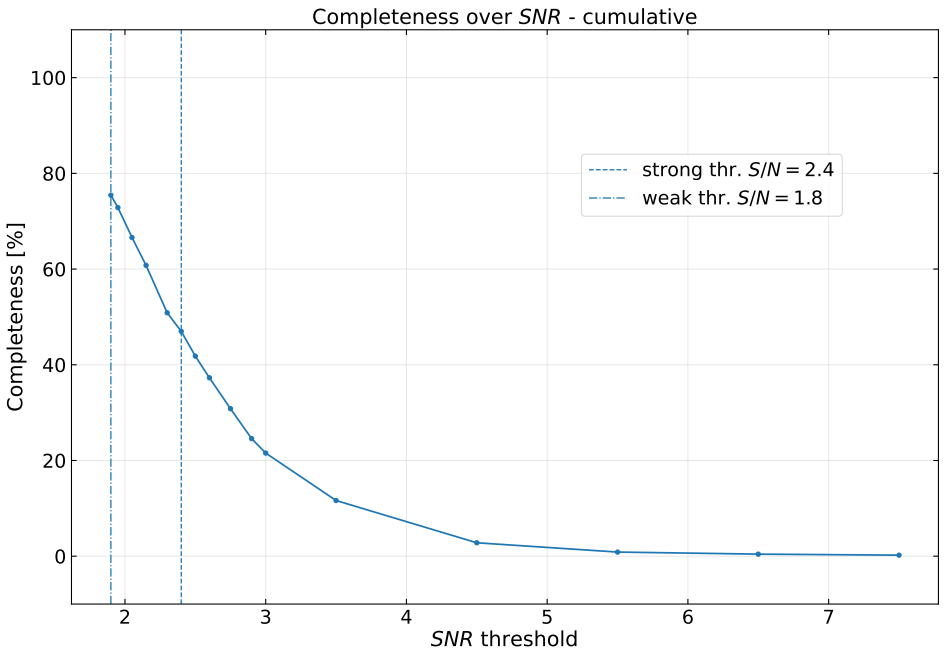


Figure 5.22 *Cumulative completeness of the sample as a function of different signal-to-noise ratio thresholds. We show the completeness of the catalog of detections for different values of SNR, from $SNR = 9.0$ down to the weak threshold of $SNR = 1.8$. With the vertical lines, we highlight the two thresholds corresponding to the weak and strong criteria for the detection.*

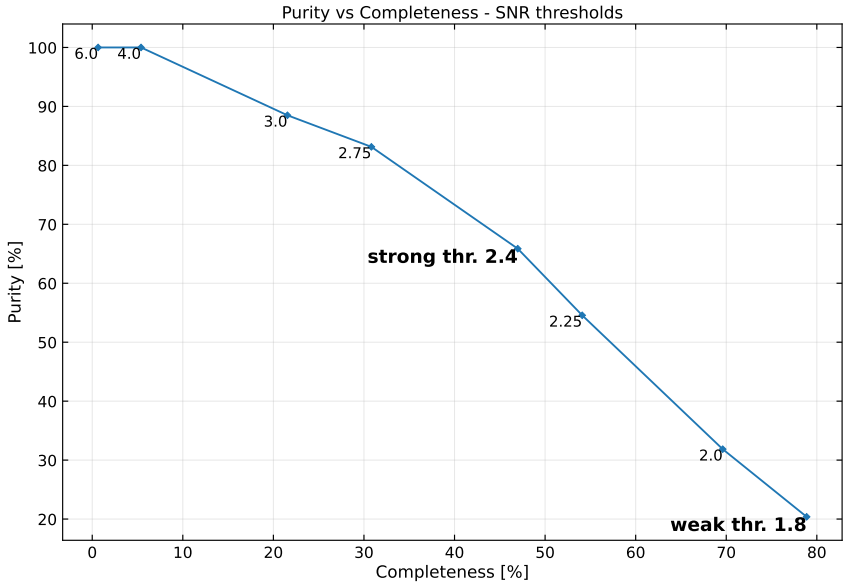


Figure 5.23 *Purity over Completeness plot for different signal-to-noise ratio thresholds. We show the purity of the catalog of detections with respect to the completeness (fraction of matched halos) for different SNR thresholds. The values of the SNR are shown as labels of the data points. As a general trend: at high SNR only a few (low completeness) but most evident objects (high purity) are detected; when lowering the SNR more objects are found (high completeness) but with the increased possibility of detecting a spurious source (low purity).*

cut the purity is high/low and the completeness is low/high. With the threshold of 2.4, the *strong* threshold, we have about 70% purity of the catalog and completeness 47%.

As the global sample completeness so far presented is not unique because the redshift-mass cut-off used to define the true cluster sample can be chosen arbitrarily, it is of great interest to investigate the purity as a function of both redshift and virial mass, M_{200}, of the matched clusters, as shown in Figure 5.24. The results are shown for both strong and relaxed SNR thresholds, the completeness is clearly higher for the sample with the lowest SNR threshold. In the same figure, the side plots show the completeness with respect to the individual parameters: on top the total completeness as a function of redshift, and on the right the total completeness as a function of mass. The 100% completeness reached in the redshift bin [0.7,0.8] has to be ignored for the very poor statistics at that redshift.

5.5.1 Random match test

To derive the number of positive matches due to pure random superposition we created 10 randomized catalogs from the original reference one presented in Section 5.4 by shuffling the RA and Dec positions of the entries. Then, we ran the matching procedure between these randomized catalogs and the catalog of detections obtained with these randomized catalogs with the catalog of detections with the same settings adopted for the unshuffled run. As a test, we verified that the matched objects have

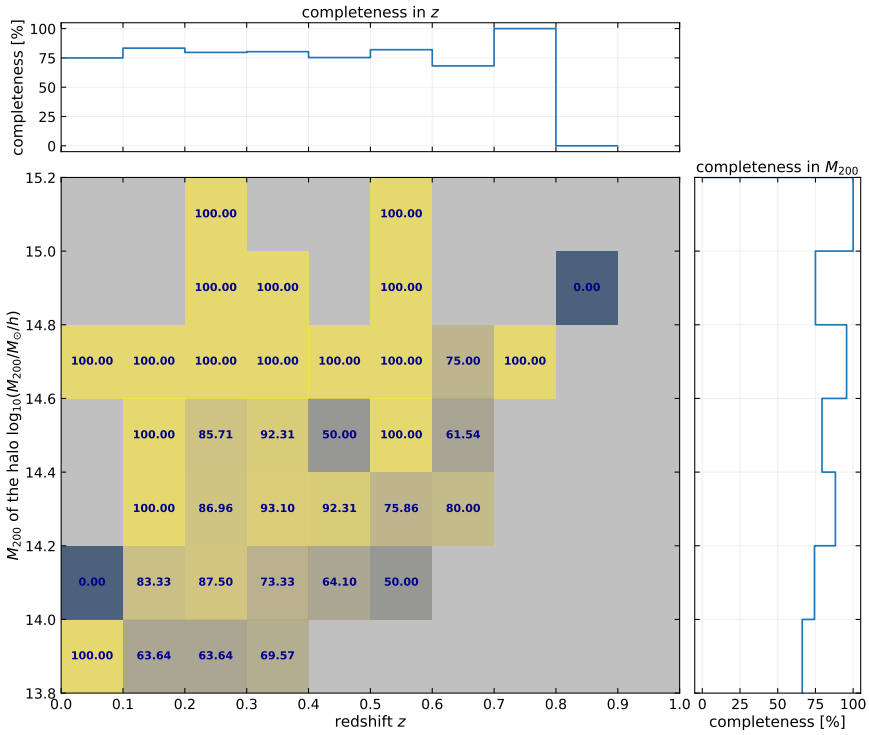


Figure 5.24 *Completeness plot in the $z - M_{200}$ plane for the catalog of detections. In **top panel** we show the completeness of the full sample inside the redshift-mass plane as a color map. We also report the value of completeness in percentage inside each bins. We use bins of 0.1 in redshift within the range [0.0,1.0] and logarithmic bins of virial mass of 0.2 dex within [13.8,15.2]. We show also as marginal plots the 1-dimensional completeness for the bins of redshift and mass. In **bottom panel** we show the same plots for the case of the strong detections subset ($SNR > 2.4$).*

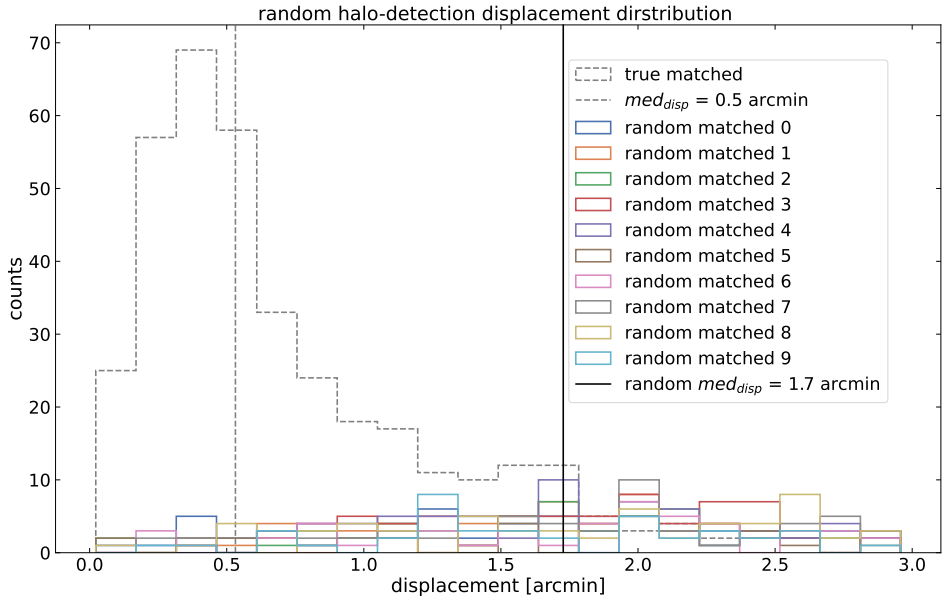


Figure 5.25 *Displacement distribution of the matched detections with the 10 random catalogs of halos. We show the displacement between the halo and the corresponding matched detection in arcmin on the projected plane. With the dashed distribution and vertical line we show the true matched halo distribution, while with the solid black line, we show the median of the 10 random displacement distributions.*

a uniform distribution in angular displacement as expected (see Figure 5.25). As a comparison, we also show the distribution which we obtained when we considered the true halo catalog.

Table 5.4 lists the average values (with 1σ error) for purity and completeness of the detections based on the random catalogs and the corresponding value obtained with the true catalog. This represents the average spurious contribution to completeness and purity due to chance projections of completeness and purity that has to change projections that has to be accounted for.

In Figure 5.26 we show the average completeness in the plane $z - M_{200}$ for the 10 random catalogs. In the same plot, we present both the case for the full sample and the high-SNR subset: the top label in each 2D bin refers to the full sample,

	random purity %	true purity %	*random* completeness %	true completeness %
$SNR > 1.8$	3.0 ± 0.4	20.4	11.7 ± 1.8	78.9
$SNR > 2.4$	4.3 ± 1.0	65.9	3.0 ± 0.5	47.0

Table 5.4 *Purity and completeness computed for the 10 randomized catalogs of halos, for both SNR thresholds. We report the average purity and completeness with their standard deviation error computed from the 10 randomized catalogs. As references, we also report purity and completeness obtained with the true halo catalog.*

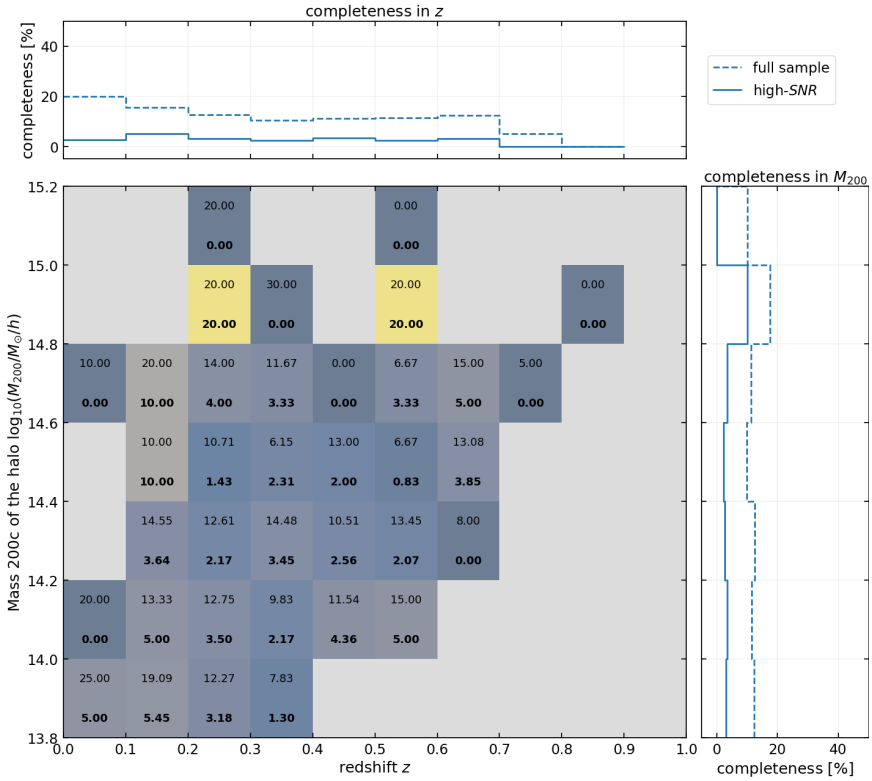


Figure 5.26 *Completeness plot in the $z - M_{200}$ plane for the two catalogs of detections (full sample and high-SNR subset) with respect to the 10 randomized catalogs of dark matter halos. We show also marginal plots with 1-D completeness with respect to redshift and virial mass. The color map is referred to the high-SNR subset of matched detections while the numbers inside the bins represent the full sample (top number) and subset (bottom bold number).*

also represented by the color map, while the bold label refers to the high-SNR subset. We have a relatively high value of random completeness ($\sim 20\%$) in two bins associated with the mass bins $14.8 - 15.0\ \log(\mathrm{M}_\odot/h)$ due to the fact that in these two bins, we have only one halo and randomly matched two times out of 10. The low statistics of halos in these bins can produce a high value of completeness in a random way.

5.6 Effect of varying the input catalog of galaxy ellipticities

As described in the previous sections, we included in our weak lensing analysis only the galaxies with $z > 0.6$ to remove in a very simple manner most of the foreground galaxies. The ellipticity of the galaxies at lower redshift is not related to the lensing signal of the majority of the clusters we aim at detecting: then they would act as noise diluting the lensing signal. In Appendix A we present the same analysis performed on this *truncated* galaxy catalog presented up to now also on

the *complete* galaxy catalog, including also lower redshift galaxies. Both strong and weak thresholds based on the *peaks* $E-B$ *method* and on the *peaks* $E-\hat{E}$ *method* have been used. In this case, the thresholds are $SNR = 1.9$ and $SNR = 2.4$, with 1189 and 287 detections, respectively.

As expected the complete catalog produces systematically fewer detections: the galaxies with $z < 0.6$, removed in the truncated catalog, are more likely to be a foreground for the halos leading to a decrease in the amplitude of the signal and an increase of the noise (see Eq. 4.16 and Eq. 4.17). In particular, we have 607 fewer detections, almost 34% less with respect to the truncated catalog, with 62 of them detected over the strong threshold.

Applying the same matching procedure, i.e. 1 h^{-1}Mpc matching radius limited to a maximum of 3 arcminutes, we obtain 323 matched objects with the true halos with the lower SNR threshold and 187 with a $SNR > 2.4$. We found 43 fewer matches ($\sim$12%) with the complete catalog than the truncated one. In Figure 5.27 we compare the distributions of matched haloes of the two data sets (truncated and complete catalog) as a function of the signal-to-noise ratio of the detection (in the top panel of Figure 5.27), the virial mass (in the central panel of Figure 5.27), and the redshift of the halos (in the bottom panel of Figure 5.27). In the figures, we show with blue lines the distributions and relative medians (vertical lines) for the truncated catalog, while in red for the complete catalog; for both cases, the dashed lines report the information for the full sample while the solid lines represent the high-SNR subset. Despite the overall decrease of matched objects for the complete catalog due to the lower number of detections, we do not see any strong difference in the distributions. We compute the median and verify that the distributions are consistent with each other. In the signal-to-noise ratio distribution, it is clear how the distribution at the low values of SNR is identical for the two catalogs, while of the 43 fewer matches in the complete catalog, 31 are associated with a $SNR > 2.4$ detections.

In Figure 5.28 we show the comparison of purity over completeness for different signal-to-noise ratio thresholds: we have in fact 79% against 70% in completeness for a fixed purity of 20%, related to the weak SNR threshold of the truncated catalog, while for a fixed purity of 66%, related to the strong SNR threshold of the truncated catalog, the completeness for the truncated catalog is 47% while is 45% for the complete catalog.

We shall now discuss the characteristics of the additional matches that the truncated catalog has with respect to the complete one. In Figure 5.29 we plot the distribution of the additional matches in the $z - M_{200}$ plane as done for the other catalogs. We have in total 56 additional matched detections based on the truncated catalog, all of which lay in low-mass regions where also the SNR are the lowest. In Figure 5.30 we show the distribution of the difference in SNR of the matched halos that are common to the detections based on the truncated and complete galaxy catalogs. On the total 310 shared matches, more than half, i.e. 66%, of the truncated galaxy catalog have higher SNR than those based on the complete one. The increase of SNR of the detections is one of the factors responsible for the increment of the sample size of matched detections for the truncated catalog. We have seen the consequences of this increment in Figure 5.23.

Finally, in Figure 5.31 we compare the completeness of the *complete* catalog and of the *truncated* one. We show the completeness in redshift and mass of the two datasets, red for the complete, and blue for the truncated. In the central plot, we

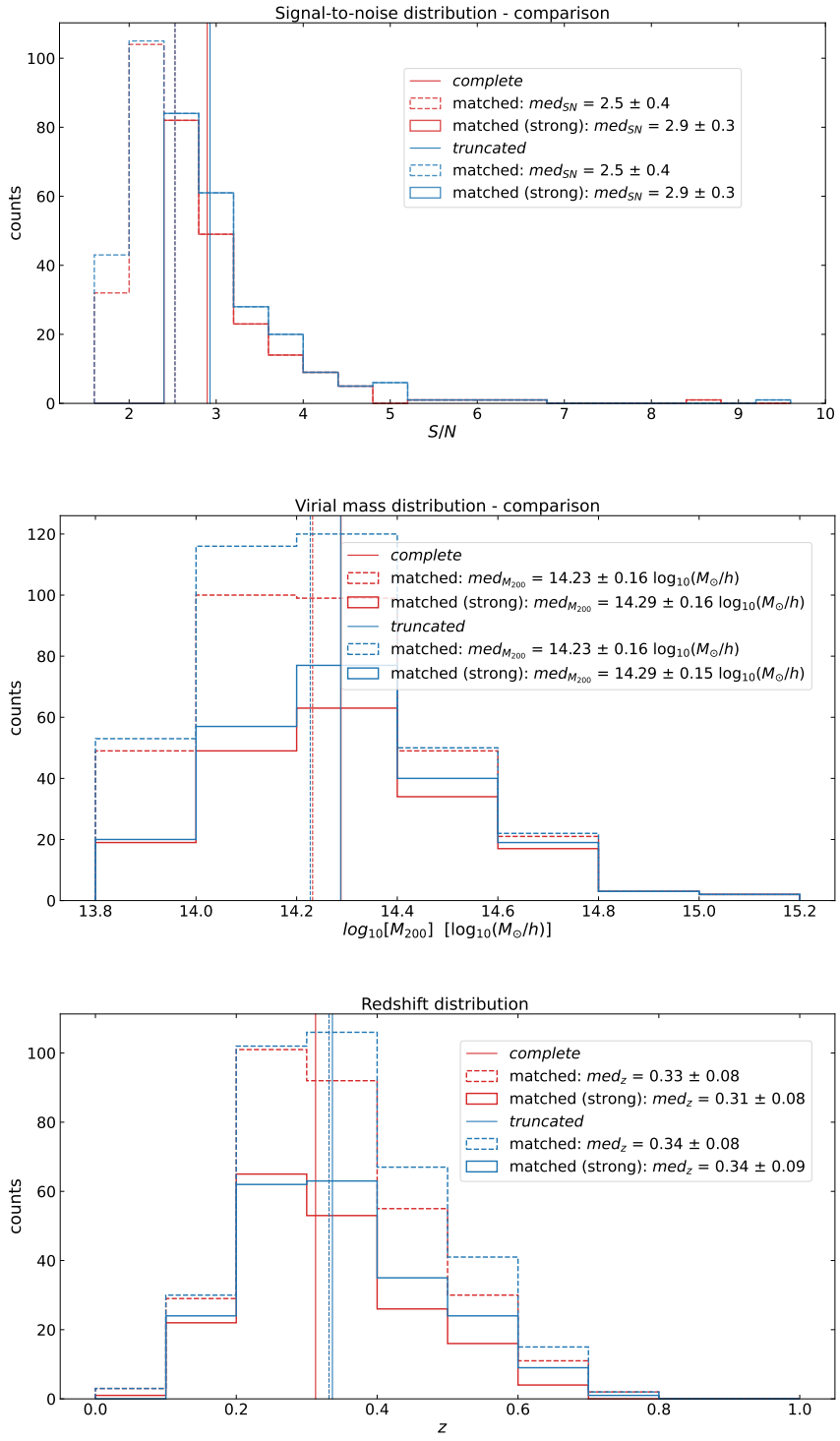


Figure 5.27 *SNR, virial mass and redshift distribution of the complete and truncated matches (full sample and high-SNR subset). In the three panels the blue lines represent the truncated catalog and the red lines the complete one; the dashed lines are related to the full sample and the solid ones to the high-SNR subset. In* ***top panel*** *we show the comparison of the SNR distribution of matched objects between the complete and truncated catalog. In* ***central panel*** *we show the comparison of the virial mass distribution of matched objects in the complete and truncated catalog using the same bins of Figure 5.24. In* ***bottom panel*** *we show the comparison of the redshift distribution of matched objects between the complete and truncated catalog using the same bins of Figure 5.24.*

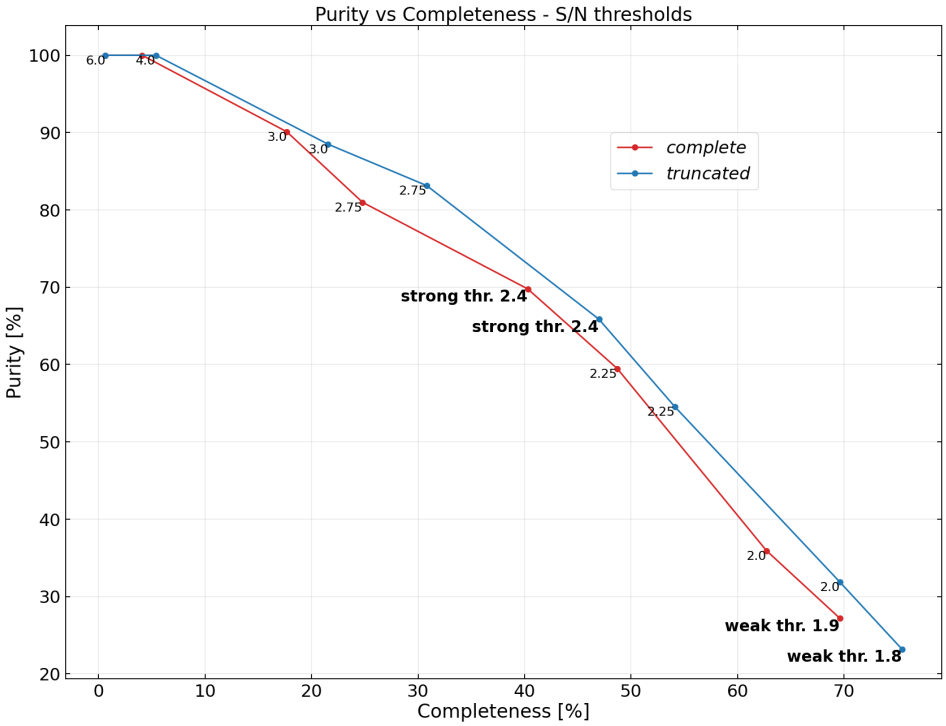


Figure 5.28 *We show the purity of the two catalogs of detections obtained using the complete and truncated catalog of galaxy ellipticities (red and blue lines, respectively) as a function of the completeness computed for different SNR thresholds. The values of the SNR are shown as labels of the data points.*

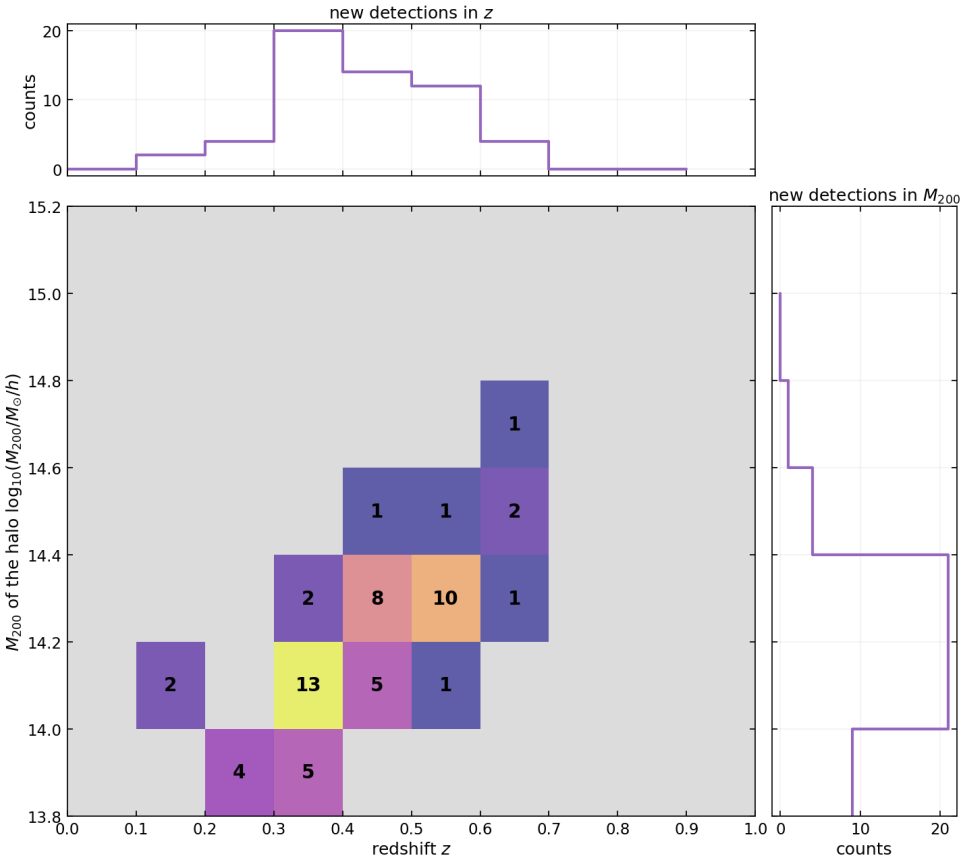


Figure 5.29 *Distribution of the additional matches that the truncated catalog has with respect to the complete one. We show the distributions of redshift and virial mass as marginal plots while in the central plot, we show the 2D distribution of the new matches in the $z - M_{200}$ plane.*

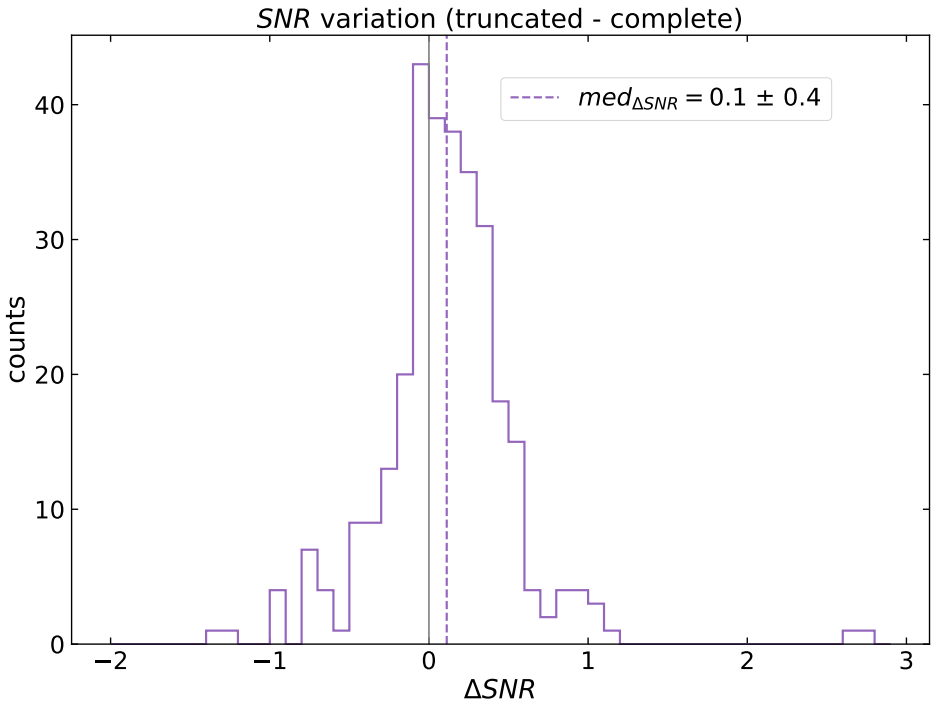


Figure 5.30 *Distribution of signal-to-noise ratio variation of halos from the truncated to the complete catalog. The gray vertical line represents* $\Delta SNR = 0$*. The dashed purple line represents the median of the signal-to-noise variation of the common detections.*

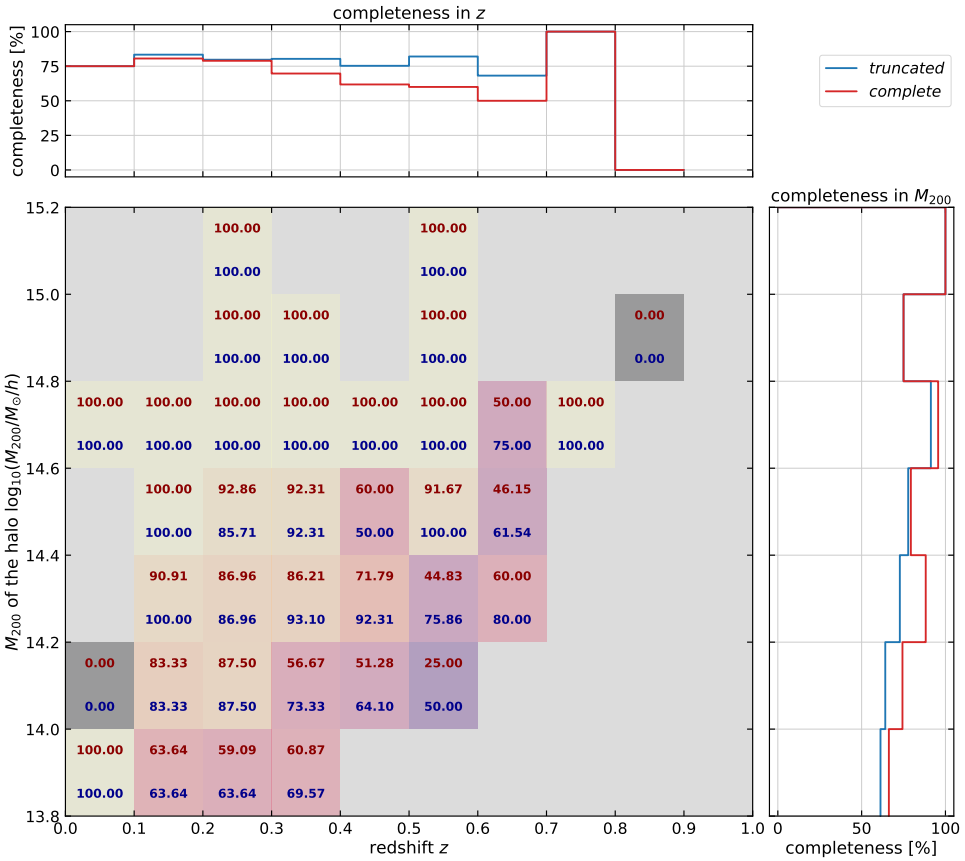


Figure 5.31 *Completeness analysis in the $z - M_{200}$ plane for both the complete (red) and truncated (blue) datasets. We show the 2-D completeness color map in the redshift-mass plane comparing the results of the complete catalog with the results of the truncated catalog in the same bin. We compare also the 1D completeness in the side plots by comparing the distribution of the truncated and complete results.*

compare the 2-dimensional completeness in the $z - M_{200}$ plane. We label all the bins with their respective completeness (red for the complete and blue for the truncated). As expected, the completeness of the *truncated* catalog is higher for smaller systems and for the higher redshift. This is not the case for a few mass-redshift bins but this is due to random fluctuations due to the poor statistics.

Chapter 6

Conclusions and future perspectives

We augmented the Adaptive Matched Identifier of Clustered Objects (AMICO) algorithm to perform weak-gravitational lensing analysis of wide-field optical survey data, i.e. catalog of galaxy ellipticity, to detect clusters of galaxies. The code implements an improved version of the "optimal" linear optimal matched filter defined in (Maturi et al., 2005). The filter assumes a mean radial profile of the halo shear pattern and a power spectrum for the noise. The noise comprehends the contribution of measurement errors and of the lensing due to the large-scale structures that act as a contaminant. The filtering process allows to distinguishing the lensing peaks of galaxy clusters from the ones of the large-scale structure and observational noise. The detection of the clusters is based on the search of maxima in signal-to-noise ratio maps returned by the filter. This last step has been improved by implementing a highly optimized *cleaning* routine: clusters are identified iteratively and at each iteration, the lensing signal of the detected cluster is removed from the map to improve the detection of possible nearby blended dark matter halos. Cleaning procedures were never applied before on a weak-lensing cluster finder with the exception of (Pace et al., 2007).

We applied the code to a catalog of simulated galaxy ellipticities distorted by the weak-gravitational lensing signal of the matter distribution simulated with the MOKA Giocoli et al. (2012) code. The mock ellipticity galaxy catalog resembles the observational properties expected for the *Euclid* mission. Therefore our analysis provides a good forecast for such mission.

We investigate four different criteria to implement the SNR threshold that needs to be set for the detection criteria which goes under the following names: *pixels* $E-B$, *pixels* $E-\hat{E}$, *peaks* $E-B$ and *peaks* $E-\hat{E}$. The first criterion is based on the study of the distribution of SNR of the pixels of the E- and B-mode amplitude maps, the threshold is found by setting a value for the "expected purity" of the sample which is approximated by the ratio of the two distributions. The second criteria follow a similar concept but instead of the fluctuations of the B-mode map, the absolute value of the negative pixels of the E-mode map is used; in this way, we account for the LSS as a source of noise. The third and fourth criteria mirror the first two but, instead of considering all pixels, they consider the SNR distribution of the peaks in the maps. We tested the performance of these approaches to set

the ideal SNR cut-off to be used in surveys and conclude that the most efficient method is the one based on positive and negative peaks of E-mode: the estimate of the sample purity is in fact more reliable as all noise contributions are accounted for.

We further refine the processing strategy by using only the galaxies with $z > 0.6$ to remove most of the non-lensed foreground sources and then compared the results with the analysis performed with the full galaxy catalog. Using the truncated dataset we detected 607 more objects. We then investigated the sample in terms of *completeness* and *purity* by matching the detections with the actual halos in the simulations. Before doing that we removed from the reference halos those objects that, because of their mass and redshift, do not produce a significant lensing signal. This was done to reduce the ambiguity given by chance projections. We found higher completeness for the truncated catalog with respect to the complete one, 47% against 45% for a fixed purity of 65%.

By investigating the purity, we found that for the threshold obtained with the *peaks* $E - \hat{E}$ *method*, the purity is close to the "expected purity", 70% against the expected 80%, in contrast with the peaks $E - B$ method we obtained a purity of 20% when aiming at 80%. The completeness has been studied as a function of the matched halos redshift and mass. In the redshift in the range [0.2,0.4], the completeness reaches almost 100% in the case of the halos with virial mass $\log_{10}(M_{200}/M_{\odot}/h) > 14.4$. As expected we reached the higher completeness where the efficiency of the gravitational lensing effect is maximum: at a position of the lens that is intermediate between the sources and the observer. Using the truncated catalog we also incremented the detection of halos with smaller masses.

We further tested the reliability of the matching procedure and estimated the number of positive matches due to pure random superposition by creating 10 catalogs with random positions based on the original reference of halos. Then we matched them with the sample of detections and derive the average completeness and purity of the randomized reference. We obtained a 3.0 ± 0.4 % random purity and 11.7 ± 1.8 % random completeness. This represents the average spurious contribution to completeness and purity due to chance projections that has to be accounted for.

We plan to improve the code by implementing an adaptive filter that would increase the flexibility of the algorithm and allow the detection of dark matter halos in a broader range of masses and redshift. We then intend to extend the code to include a tomographic approach to exploit the available photometric redshifts. This would further improve the reliability of the detection. The ultimate goal is to build an optimized routine for the weak-lensing detection of clusters complementary to the photometric detection algorithms such as the original AMICO employed in the data analysis pipeline of Euclid, KiDS, and J-PAS.

The AMICO photometric algorithm in combination with the new AMICO-WL represents an extremely versatile and powerful tool for the detection of clusters of all kinds. The plan is to apply the code to other wide-field survey data like KiDS, CHFTLens, DES, and in the future LSST from the Vera C. Rubin Observatory besides Euclid.

Appendix A

Analysis of the complete ellipticity catalog

In this appendix, we show the analogous plots already shown in Chapter 5, but for the case of the *complete* catalog of galaxy ellipticity. The direct comparison of the most relevant results of the two datasets is in Section 5.6.

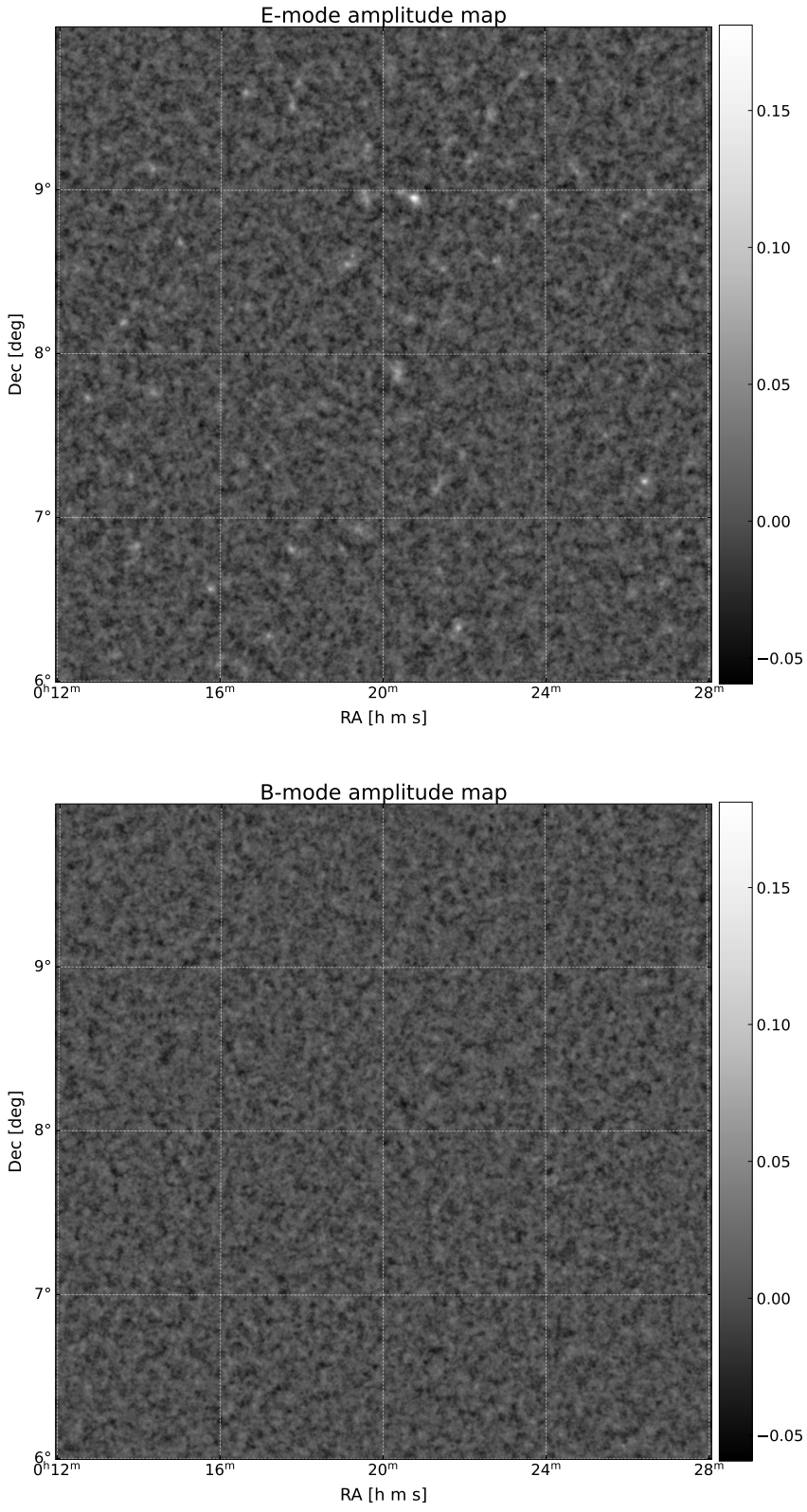


Figure A.1 *Zoom in of the output maps in the region RA∈[3.0,7.0] deg and Dec∈[6.0,10.0]: the E-mode and the B-mode are shown in the top and bottom panels, respectively. They are plotted with the same color scale. In the E-mode map, it is possible to see the typical signatures of the presence of halos: white high-value spots. These are not present in the B-mode map, that only contains the noise from galaxies.*

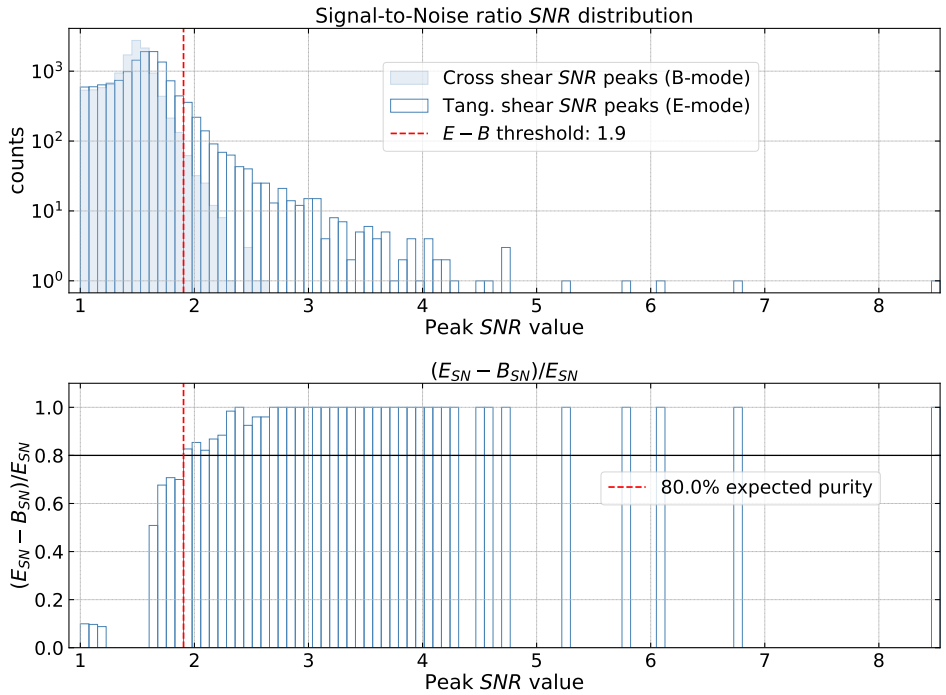


Figure A.2 *Peaks $E-B$ method: comparison of the SNR distribution of the peaks of the E-mode and B-mode amplitude maps, and difference of the histograms relative to the truncated catalog. In the* ***top panel*** *we show the SNR distribution of the peaks of the E-mode map (empty histogram) and the B-mode map (full histogram). The peaks of the B-mode map, caused by noise only, reach SNR values that are quite lower with respect to the E-mode peaks. In the* ***bottom panel*** *we have the difference of the counts in each bin of the E-mode map minus the counts of the same bin of the B-mode map, normalized over the E-mode counts. We define the threshold (red vertical line) with a given expected purity, 80%, as the SNR value in which the E-mode map counts become the 80% of the total (horizontal line); we plot the threshold also in the top figure as a vertical red line.*

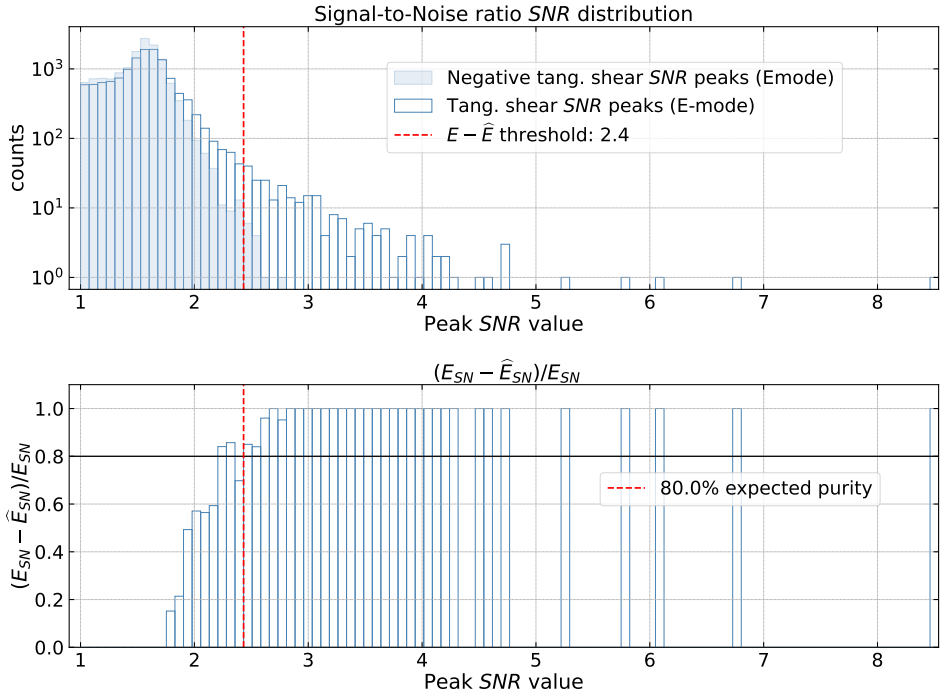


Figure A.3 *Peaks $E-\hat{E}$ method: comparison of the SNR distribution of the positive peaks and the negative peaks of the E-mode amplitude map ($\hat{E}$), and difference of the histograms, relative to the truncated catalog. In* ***top panel*** *we show the SNR distribution of the positive peaks of the E-mode map (empty histogram) and the negative peaks in absolute value, $\hat{E}$ (full histogram). The negative peaks are caused by the noise and by the negative signal of the lensing effect that is not related to galaxy clusters. In the* ***bottom panel*** *we have the difference of the counts in each bin of the positive and negative peaks of the E-mode. We define the threshold (red vertical line) with a given expected purity, 80%, as the SNR value in which the positive pixels map counts become the 80% of the total (horizontal line); we plot the threshold also in the top figure as a vertical red line.*

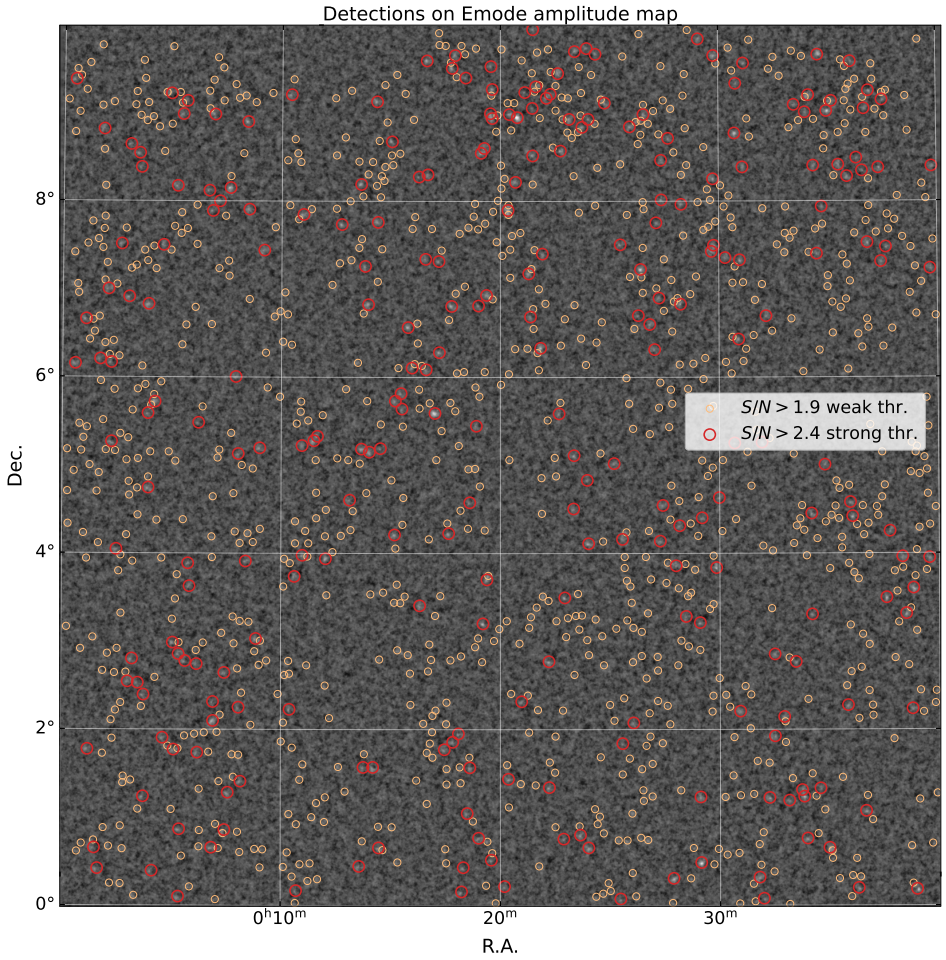


Figure A.4 *Map of the detections from the truncated catalog. We show the positions of the detections on the amplitude map where. We use the bigger blue circles for the detections with $SNR > 2.4$, the strong threshold, while the smaller light-blue circles correspond to the detections with $SNR > 1.9$, the weak threshold.*

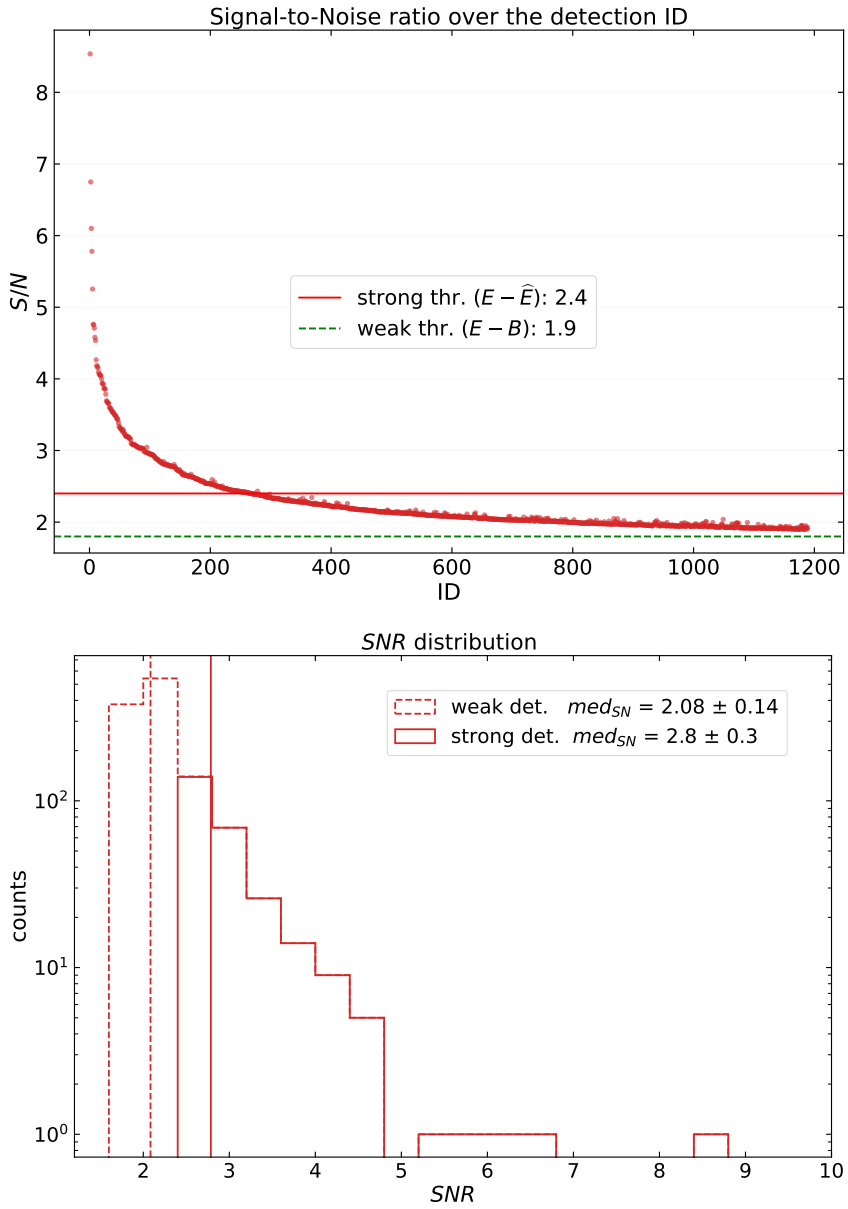


Figure A.5 *Signal-to-noise ratio analysis of the catalog of detections. In* ***top panel*** *we show the SNR of the detections as a function of their ID in the catalog, which can be used to follow the different steps of the detection procedure which starts from the higher SNR that is ∼9.0 and then rapidly decreases reaching $SNR \sim 3.0$ around ID∼100. Below the strong threshold (solid red horizontal line) the rare jumps due to the cleaning become more common. The green dashed line represents the weak threshold. In* ***bottom panel*** *we show the histogram of the SNR distribution of the detection catalog, the counts are in logarithmic scale; with the solid line we enhance the distribution of the peaks with $SNR > 2.4$, the subset found with the strong threshold, while the dashed line is the full sample of detections. The vertical lines stand for the median of the distributions.*

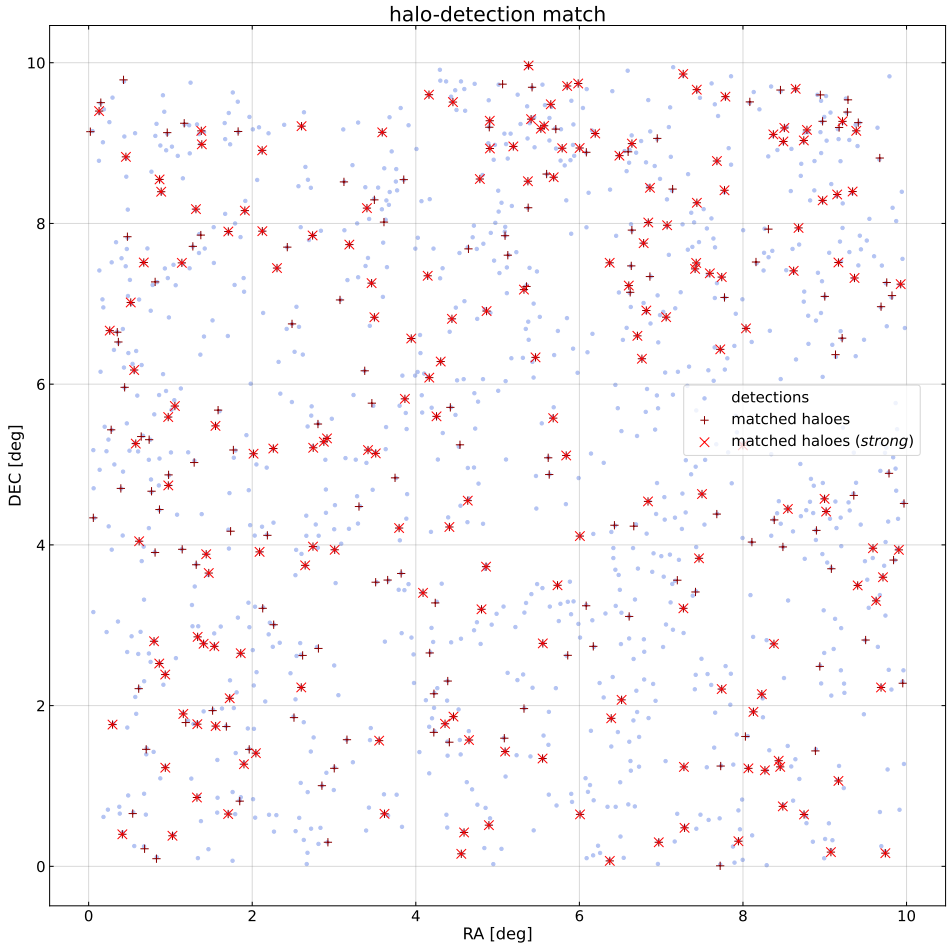


Figure A.6 *Distribution of the matched detections. We show in light blue all the detections of the complete catalog. The red plus symbol stands for the matched halo with detection in the weak threshold catalog ($SNR > 1.9$) while the bigger red cross symbol is associated only with the detections in the strong threshold catalog ($SNR > 2.4$).*

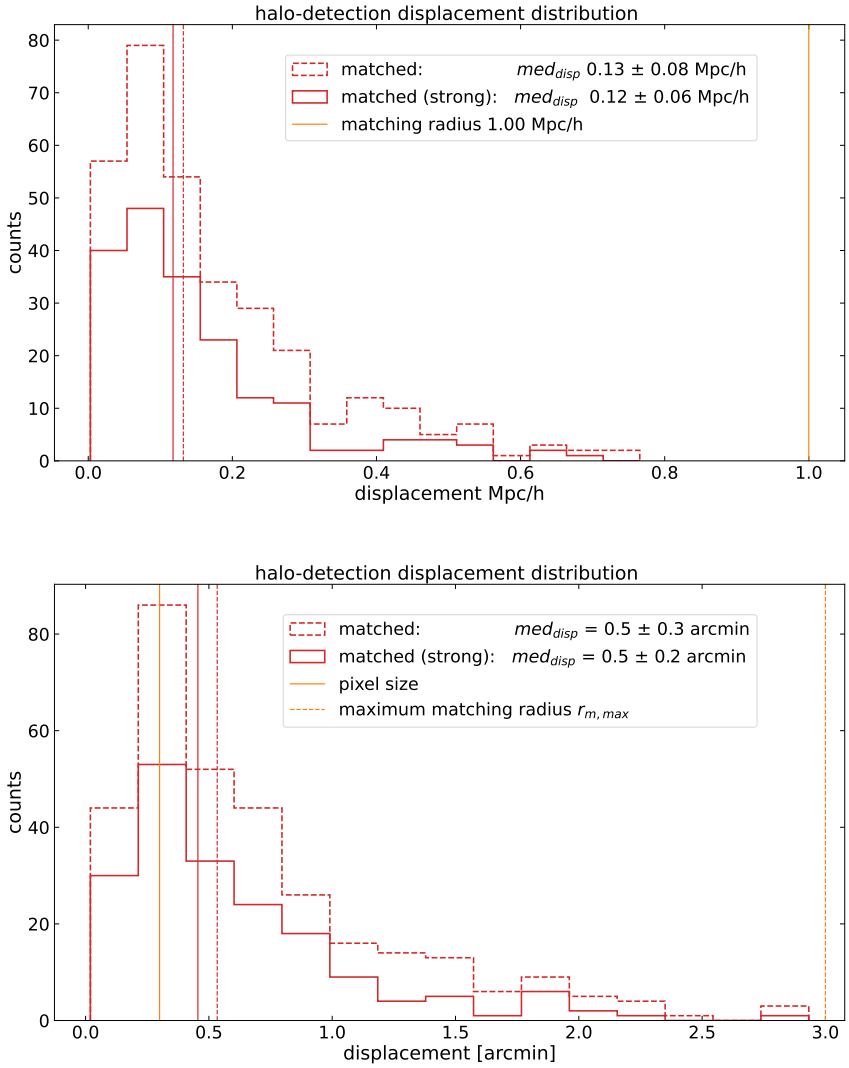


Figure A.7 *Halo-Detection displacement distribution. In the* ***top panel*** *we show the distribution of the spatial distance between the position of the halo and the position of the detection that matches the halo. We expect the position of the halo and the corresponding detection to be slightly different due to the use of grids in AMICO-WL. We show the medians of the distributions (vertical lines) and the 1 h^{-1}Mpc matching radius (orange vertical line). In the* ***bottom panel*** *we show the distribution of the displacement in the projected plane. The vertical orange line shows the dimension of the pixel used in AMICO-WL map creation. In both figures the dashed line represents the matched detections using the weak threshold while the solid represents the subset of matched detections using the strong threshold.*

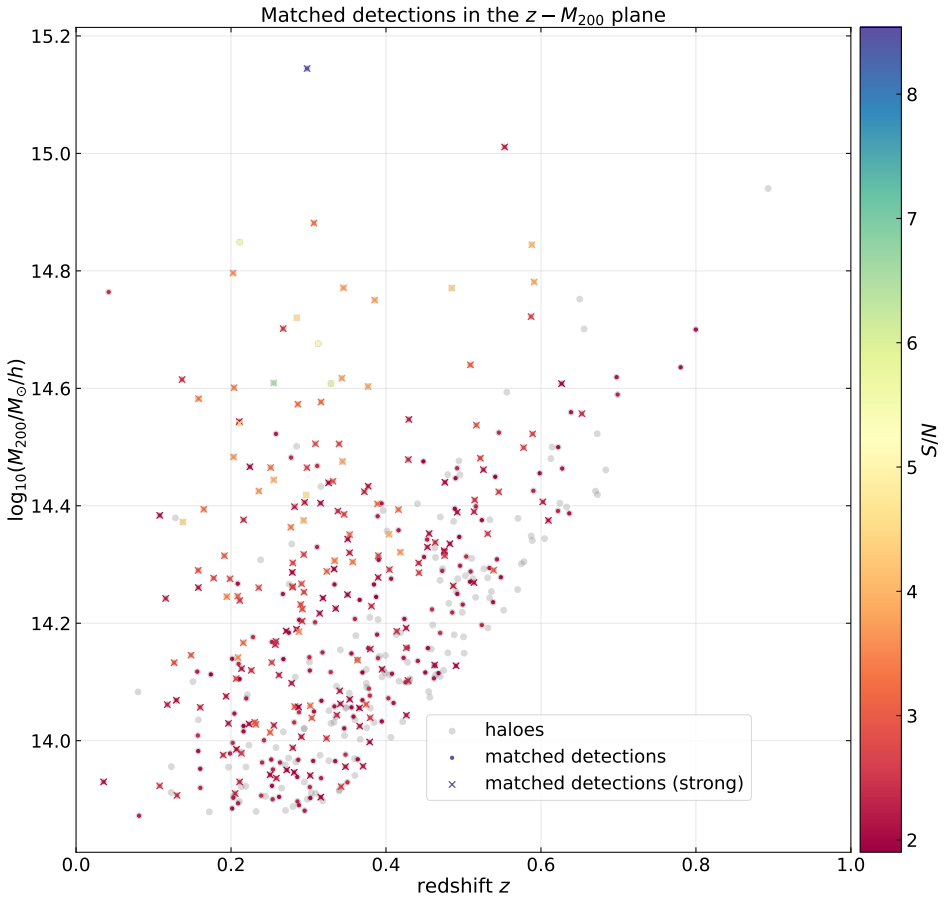


Figure A.8 *Distribution of the halos as a function of Mass and Redshift. Grey dots represent all the 464 simulated halos. The colored dots are the full sample of matched halos, while the ones in the $SNR > 2.4$ subset are highlighted with a cross symbol. The color scale of the dots and cross symbols represents the SNR of the detections. The halos in the regions with high mass, $\log_{10} M_{200}/M_\odot)/h > 14.3$, and redshift $\approx 0.2 - 0.4$ are mostly matched and with high-SNR detections.*

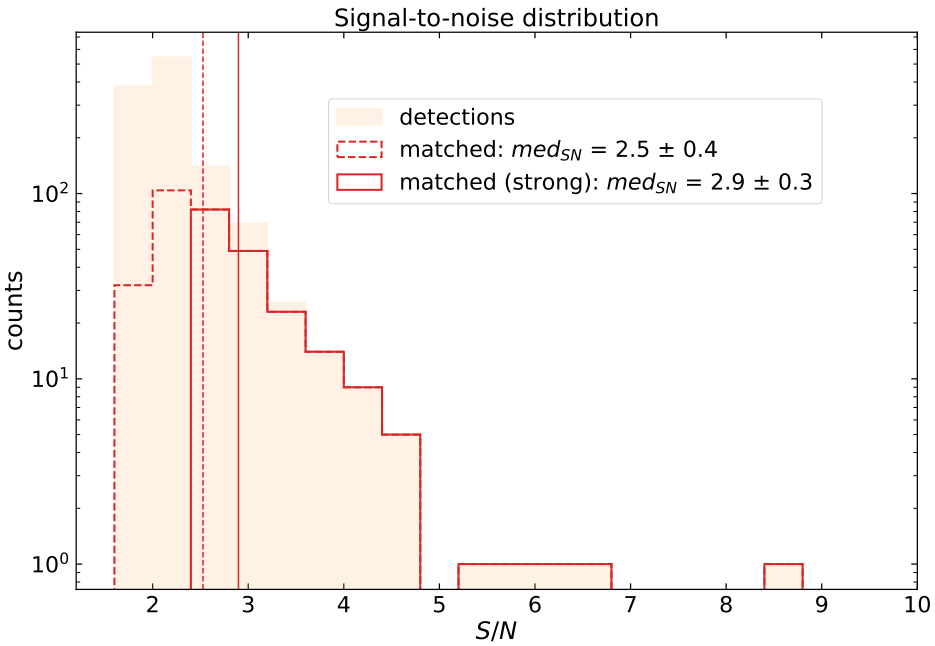


Figure A.9 *Signal-to-noise ratio distribution of the matched detections. We show the distribution of the SNR for the weak threshold detections (dashed histogram) and the subset of the strong threshold detections (solid histogram); we show the medians with vertical lines of the same style. The background shaded histogram represents the SNR distribution of the whole detection catalog. It is clear that most of the un-matched detections have a low signal-to-noise ratio.*

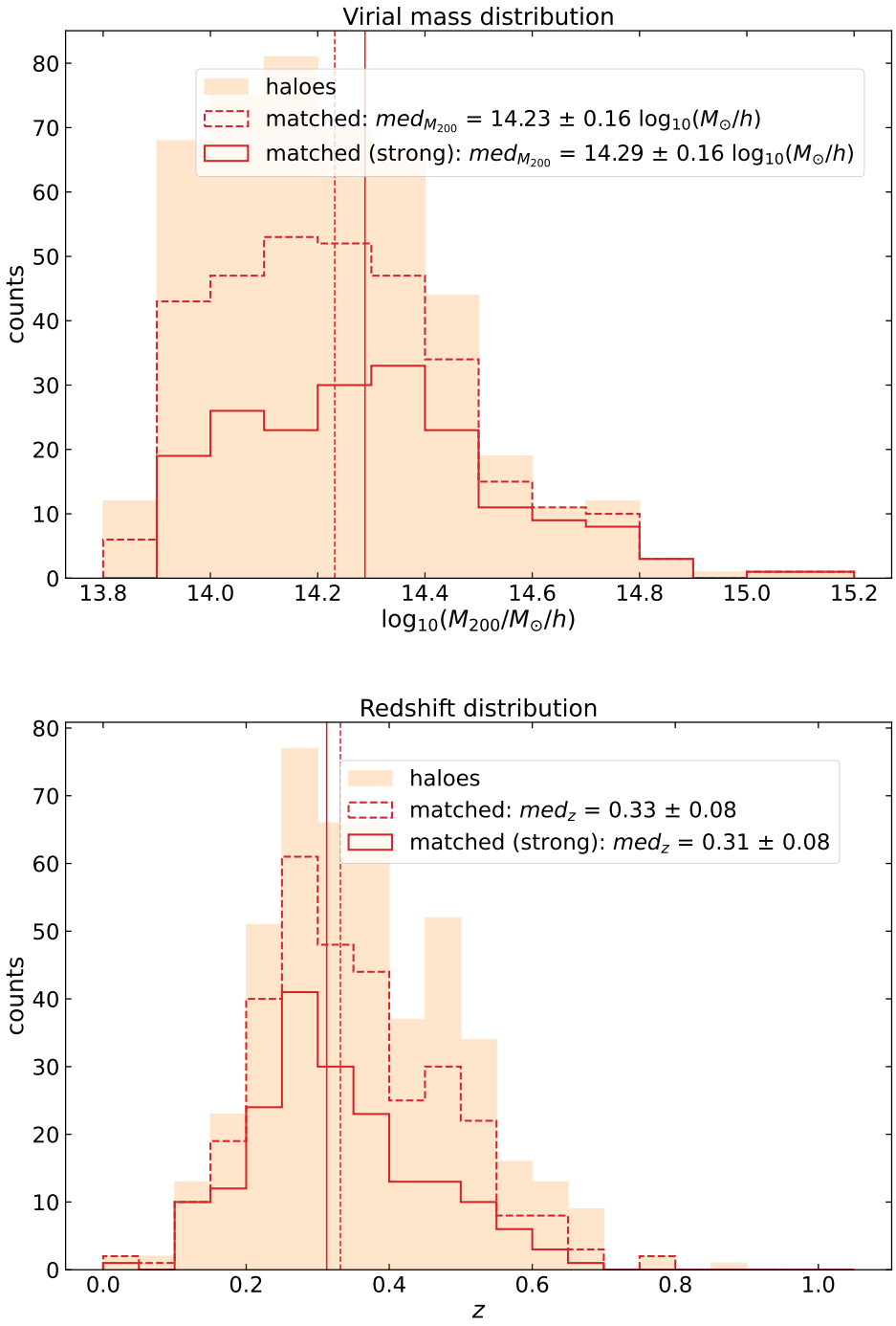


Figure A.10 *In **top panel** we show the distribution of virial mass (in logarithm) while in **bottom panel** we have the redshift distribution for the matched halos. The background shaded orange histogram is the distribution of the total 464 simulated dark matter halos. In both panels, the dashed line refers to the full sample, and the solid line to the high-SNR subset.*

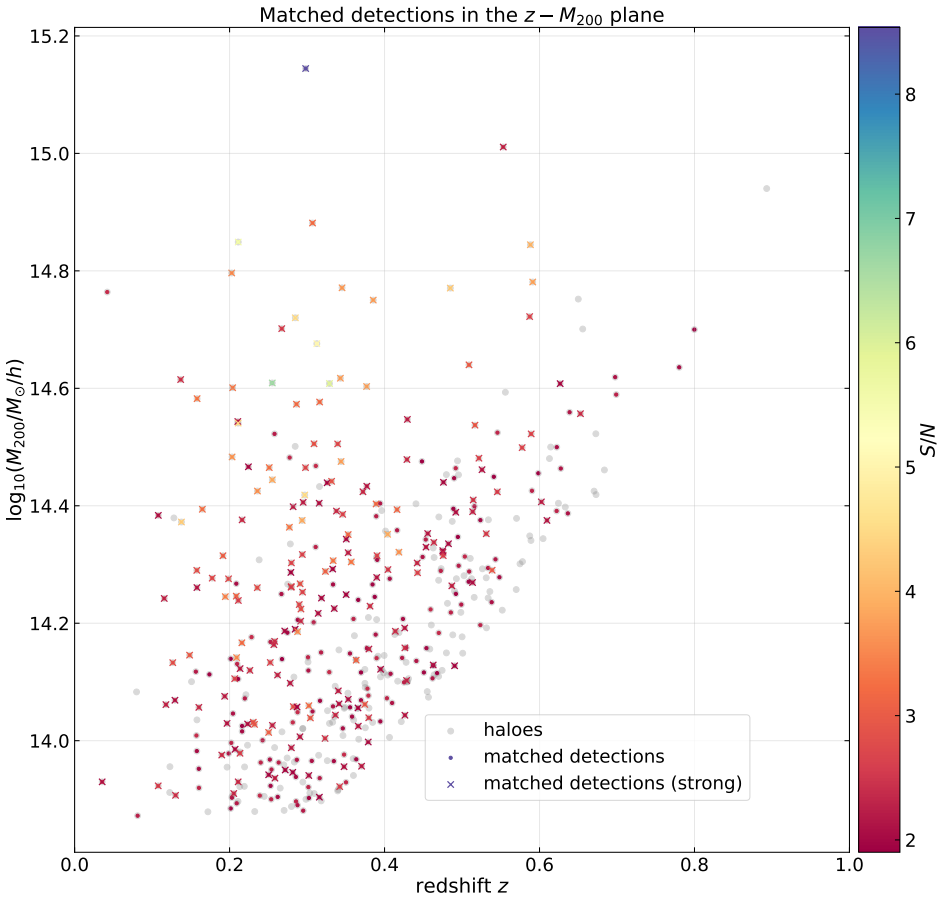


Figure A.11 *Distribution of the halos as a function of Mass and Redshift. Grey dots are all the 464 simulated halos. The colored dots are the full sample of matched halos, while the ones in the $SNR > 2.4$ subset are highlighted with a cross. The color scale of the dots and cross symbol represent the SNR of the detections. The halos in the regions with high mass, $M_{200} > 14.3 \log_{10}(h^{-1} M_\odot)$, and redshift $\approx 0.2 - 0.4$ are mostly matched and with high-SNR detections.*

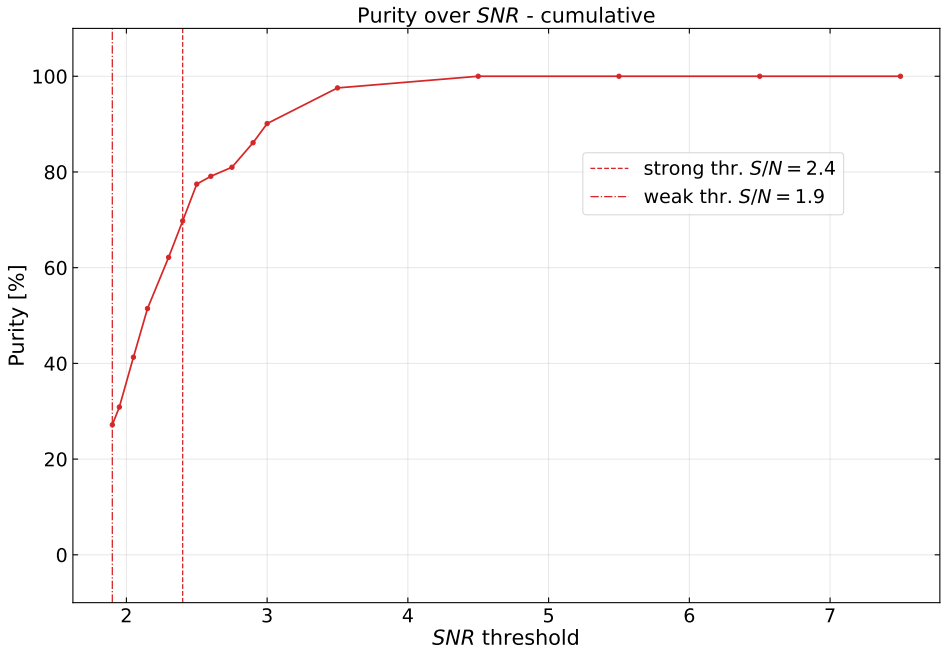


Figure A.12 *Cumulative purity of the sample over different signal-to-noise ratio thresholds. We show the purity of the catalog of detections over different values of SNR, from $SNR = 8.0$ down to the weak threshold of $SNR = 1.8$. With the vertical lines, we highlight the two thresholds corresponding to the weak and strong criteria for the detection.*

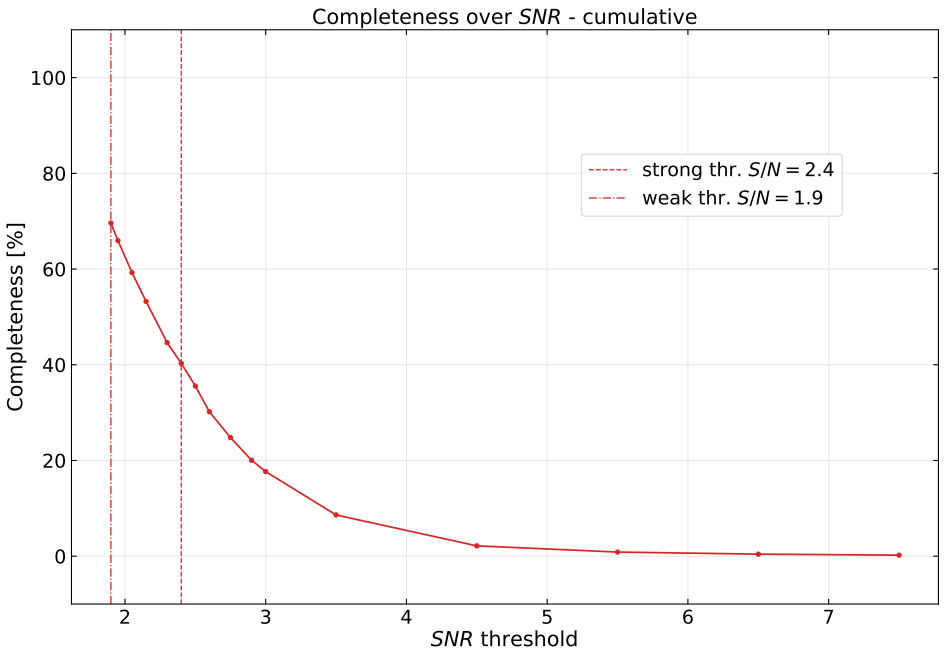


Figure A.13 *Cumulative completeness of the sample as a function of different signal-to-noise ratio thresholds. We show the completeness of the catalog of detections for different values of SNR, from $SNR = 9.0$ down to the weak threshold of $SNR = 1.8$. With the vertical lines, we highlight the two thresholds corresponding to the weak and strong criteria for the detection.*

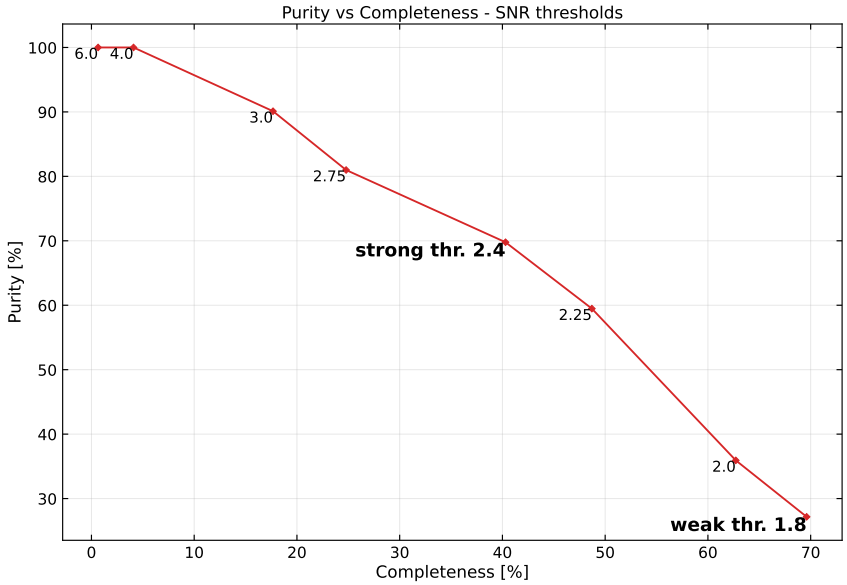


Figure A.14 *Purity over Completeness plot for different signal-to-noise ratio thresholds. We show the purity of the catalog of detections with respect to the completeness (fraction of matched halos) for different SNR thresholds. The values of the SNR are shown as labels of the data points. As a general trend: at high SNR only a few (low completeness) but most evident objects (high purity) are detected; when lowering the SNR more objects are found (high completeness) but with the increased possibility of detecting a spurious source (low purity).*

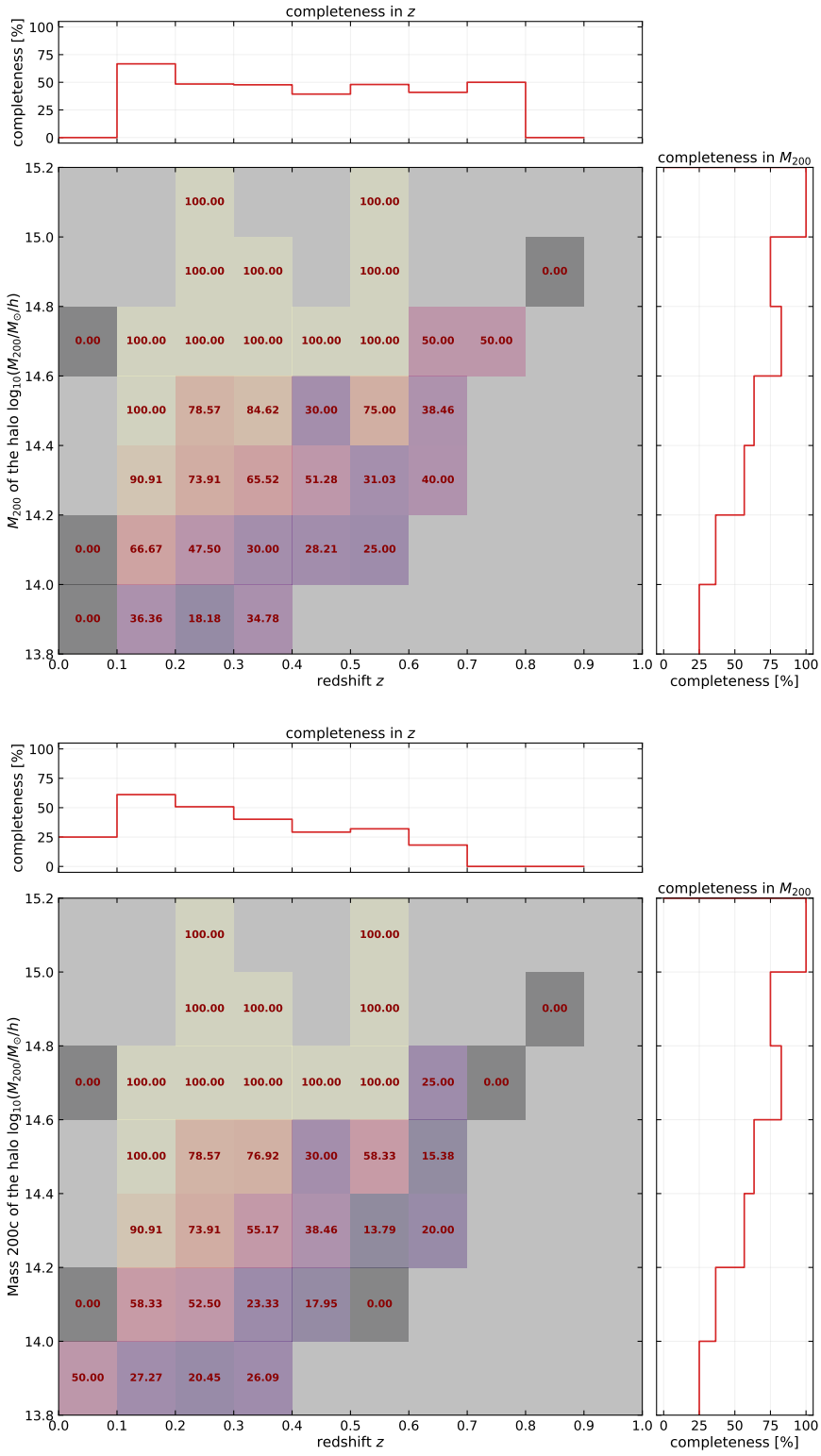


Figure A.15 *Completeness plot in the $z - M_{200}$ plane for the catalog of detections. In **top panel** we show the completeness of the full sample inside the redshift-mass plane as a color map. We also report the value of completeness in percentage inside each bins. We use bins of 0.1 in redshift within the range [0.0,1.0] and logarithmic bins of virial mass of 0.2 dex within [13.8,15.2]. We show also as marginal plots the 1-dimensional completeness for the bins of redshift and mass. In **bottom panel** we show the same plots for the case of the strong detections subset ($SNR > 2.4$).*

Bibliography

M. H. Abdullah, A. Klypin, and G. Wilson. Cosmological constraints on ωm and σ8 from cluster abundances using the galwcat19 optical-spectroscopic sdss catalog. *The Astrophysical Journal*, 901(2):90, 2020.

R. Adam, M. Vannier, S. Maurogordato, A. Biviano, C. Adami, B. Ascaso, F. Bellagamba, C. Benoist, A. Cappi, A. Díaz-Sánchez, et al. Euclid preparation-iii. galaxy cluster detection in the wide photometric survey, performance and algorithm selection. *Astronomy & Astrophysics*, 627:A23, 2019.

N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. Banday, R. Barreiro, N. Bartolo, S. Basak, et al. Planck 2018 results-vi. cosmological parameters. *Astronomy & Astrophysics*, 641:A6, 2020.

S. W. Allen, A. E. Evrard, and A. B. Mantz. Cosmological parameters from observations of galaxy clusters. *Annual Review of Astronomy and Astrophysics*, 49:409–470, 2011.

M. Bartelmann. Gravitational lensing. *Classical and Quantum Gravity*, 27(23): 233001, 2010.

M. Bartelmann and M. Maturi. Weak gravitational lensing. *arXiv preprint arXiv:1612.06535*, 2016.

M. Bartelmann and P. Schneider. Weak gravitational lensing. *Physics Reports*, 340 (4-5):291–472, 2001.

F. Bellagamba, M. Maturi, T. Hamana, M. Meneghetti, S. Miyazaki, and L. Moscardini. Optimal filtering of optical and weak lensing data to search for galaxy clusters: application to the cosmos field. *Monthly Notices of the Royal Astronomical Society*, 413(2):1145–1157, 2011.

F. Bellagamba, M. Roncarelli, M. Maturi, and L. Moscardini. Amico: optimized detection of galaxy clusters in photometric surveys. *Monthly Notices of the Royal Astronomical Society*, 473(4):5221–5236, 2018.

F. Bellagamba, M. Sereno, M. Roncarelli, M. Maturi, M. Radovich, S. Bardelli, E. Puddu, L. Moscardini, F. Getman, H. Hildebrandt, et al. Amico galaxy clusters in kids-dr3: weak lensing mass calibration. *Monthly Notices of the Royal Astronomical Society*, 484(2):1598–1615, 2019.

L. Bleem, B. Stalder, T. De Haan, K. Aird, S. Allen, D. Applegate, M. L. Ashby, M. Bautz, M. Bayliss, B. Benson, et al. Galaxy clusters discovered via the sunyaev–zel'dovich effect in the 2500-square-degree spt-sz survey. *The Astrophysical Journal Supplement Series*, 216(2):27, 2015.

R. G. Bower, J. Lucey, and R. S. Ellis. Precision photometry of early-type galaxies in the coma and virgo clusters: a test of the universality of the colour–magnitude relation–i. the data. *Monthly Notices of the Royal Astronomical Society*, 254(4): 589–600, 1992.

S. M. Carroll. An introduction to general relativity: spacetime and geometry. *Addison Wesley*, 101:102, 2004.

R. G. Crittenden, P. Natarajan, U.-L. Pen, and T. Theuns. Discriminating weak lensing from intrinsic spin correlations using the curl-gradient decomposition. *The Astrophysical Journal*, 568(1):20–27, 3 2002. doi: 10.1086/338838. URL https://doi.org/10.1086%2F338838.

V. Desjacques, D. Jeong, and F. Schmidt. Large-scale galaxy bias. *Physics reports*, 733:1–193, 2018.

F. W. Dyson, A. S. Eddington, and C. Davidson. Ix. a determination of the deflection of light by the sun's gravitational field, from observations made at the total eclipse of may 29, 1919. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 220(571-581): 291–333, 1920.

A. Einstein. Die grundlage der allgemeinen relativitätstheorie annalen der physik, 49. *Reprinted in English translation in The Principle of Relativity.(1952) Dover Publications Inc, New York*, 1916.

A. Einstein and W. De Sitter. On the relation between the expansion and the mean density of the universe. *Proceedings of the National Academy of Sciences*, 18(3): 213–214, 1932.

P. Fischer. A new weak-lensing analysis of ms 1224.7+ 2007. *The Astronomical Journal*, 117(5):2024, 1999.

D. Fixsen. The temperature of the cosmic microwave background. *The Astrophysical Journal*, 707(2):916, 2009.

W. L. Freedman and B. F. Madore. The hubble constant. *Annual Review of Astronomy and Astrophysics*, 48:673–710, 2010.

A. Friedmann. Über die krümmung des raumes. *Z. Phys.*, 10:377–386, 1922.

A. Friedmann. Über die möglichkeit einer welt mit konstanter negativer krümmung des raumes. *Zeitschrift für Physik*, 21(1):326–332, 1924.

C. Giocoli, M. Meneghetti, M. Bartelmann, L. Moscardini, and M. Boldrin. Moka: a new tool for strong lensing studies. *Monthly Notices of the Royal Astronomical Society*, 421(4):3343–3355, 2012.

C. Giocoli, R. B. Metcalf, M. Baldi, M. Meneghetti, L. Moscardini, and M. Petkova. Disentangling dark sector models using weak lensing statistics. *Monthly Notices of the Royal Astronomical Society*, 452(3):2757–2772, 2015.

C. Giocoli, M. Baldi, and L. Moscardini. Weak lensing light-cones in modified gravity simulations with and without massive neutrinos. *Monthly Notices of the Royal Astronomical Society*, 481(2):2813–2828, 2018.

C. Giocoli, F. Marulli, L. Moscardini, M. Sereno, A. Veropalumbo, L. Gigante, M. Maturi, M. Radovich, F. Bellagamba, M. Roncarelli, et al. Amico galaxy clusters in kids-dr3: cosmological constraints from large-scale stacked weak lensing profiles. *arXiv preprint arXiv:2103.05653*, 2021.

A. Gonzalez. in building the euclid cluster survey - scientific program, proceedings of a conference held july 6-11 2014 at the sexten center for astrophysics. 2019.

M. E. Gray, R. S. Ellis, J. R. Lewis, R. G. McMahon, and A. E. Firth. Infrared constraints on the dark mass concentration observed in the cluster abell 1942. *Monthly Notices of the Royal Astronomical Society*, 325(1):111–118, 2001.

E. R. Harrison. Fluctuations at the threshold of classical cosmology. *Physical review D*, 1(10):2726, 1970.

S. Hilbert, A. Barreira, G. Fabbian, P. Fosalba, C. Giocoli, S. Bose, M. Calabrese, C. Carbone, C. T. Davies, B. Li, et al. The accuracy of weak lensing simulations. *Monthly Notices of the Royal Astronomical Society*, 493(1):305–319, 2020.

H. Hoekstra. Weak gravitational lensing. In *New Horizons for Observational Cosmology*, pages 59–100. IOS Press, 2014.

E. Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the national academy of sciences*, 15(3):168–173, 1929.

N. Kaiser. *On the spatial correlations of Abell clusters*, volume 284. 1984.

N. Kaiser and G. Squires. Mapping the dark matter with weak gravitational lensing. *The Astrophysical Journal*, 404:441–450, 1993.

R. Laureijs, J. Amiaux, S. Arduini, J.-L. Augueres, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet, et al. Euclid definition study report. *arXiv preprint arXiv:1110.3193*, 2011.

G. Lemaître. Un univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Société Scientifique de Bruxelles, A47, p. 49-59*, 47:49–59, 1927.

G. Lesci, F. Marulli, L. Moscardini, M. Sereno, A. Veropalumbo, M. Maturi, C. Giocoli, M. Radovich, F. Bellagamba, M. Roncarelli, et al. Amico galaxy clusters in kids-dr3: Cosmological constraints from counts and stacked weak lensing. *Astronomy & Astrophysics*, 659:A88, 2022.

A. Mantz, S. W. Allen, H. Ebeling, D. Rapetti, and A. Drlica-Wagner. The observed growth of massive galaxy clusters–ii. x-ray scaling relations. *Monthly Notices of the Royal Astronomical Society*, 406(3):1773–1795, 2010.

F. Marulli, A. Veropalumbo, and M. Moresco. Cosmobolognalib: C++ libraries for cosmological calculations. *Astronomy and Computing*, 14:35–42, 2016.

M. Maturi, M. Meneghetti, M. Bartelmann, K. Dolag, and L. Moscardini. An optimal filter for the detection of galaxy clusters through weak lensing. *Astronomy & Astrophysics*, 442(3):851–860, 2005.

M. Maturi, M. Schirmer, M. Meneghetti, M. Bartelmann, and L. Moscardini. Searching dark-matter halos in the gabods survey. *Astronomy & Astrophysics*, 462(2):473–479, 2007.

M. Maturi, F. Bellagamba, M. Radovich, M. Roncarelli, M. Sereno, L. Moscardini, S. Bardelli, and E. Puddu. Amico galaxy clusters in kids-dr3: sample properties and selection function. *Monthly Notices of the Royal Astronomical Society*, 485 (1):498–512, 2019.

M. Meneghetti. *Introduction to Gravitational Lensing: With Python Examples*, volume 956. Springer Nature, 2021.

M. Meneghetti, M. Bartelmann, and L. Moscardini. Cluster cross-sections for strong lensing: analytic and numerical lens models. *Monthly Notices of the Royal Astronomical Society*, 340(1):105–114, 2003.

J. F. Navarro. The structure of cold dark matter halos. In *Symposium-international astronomical union*, volume 171, pages 255–258. Cambridge University Press, 1996.

J. F. Navarro, E. Hayashi, C. Power, A. Jenkins, C. S. Frenk, S. D. White, V. Springel, J. Stadel, and T. R. Quinn. The inner structure of λcdm haloes–iii. universality and asymptotic slopes. *Monthly Notices of the Royal Astronomical Society*, 349 (3):1039–1051, 2004.

N. Okabe and K. Umetsu. Subaru weak lensing study of seven merging clusters: distributions of mass and baryons. *Publications of the Astronomical Society of Japan*, 60(2):345–375, 2008.

F. Pace, M. Maturi, M. Meneghetti, M. Bartelmann, L. Moscardini, and K. Dolag. Testing the reliability of weak lensing cluster detections. *Astronomy & Astrophysics*, 471(3):731–742, 2007.

A. A. Penzias and R. W. Wilson. A measurement of excess antenna temperature at 4080 mc/s. *Astrophysical Journal, vol. 142, p. 419-421*, 142:419–421, 1965.

S. Perlmutter, G. Aldering, G. Goldhaber, R. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, et al. Measurements of ω and λ from 42 high-redshift supernovae. *The Astrophysical Journal*, 517(2):565, 1999.

W. H. Press and P. Schechter. Formation of galaxies and clusters of galaxies by self-similar gravitational condensation. *The Astrophysical Journal*, 187:425–438, 1974.

A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The astronomical journal*, 116(3):1009, 1998.

H. P. Robertson. Kinematics and world-structure. *The Astrophysical Journal*, 82: 284, 1935.

M. Roncarelli, L. Moscardini, S. Borgani, and K. Dolag. The sunyaev–zel'dovich effects from a cosmological hydrodynamical simulation: large-scale properties and correlation with the soft x-ray signal. *Monthly Notices of the Royal Astronomical Society*, 378(4):1259–1269, 2007.

P. Rosati, S. Borgani, and C. Norman. The evolution of x-ray clusters of galaxies. *Annual Review of Astronomy and Astrophysics*, 40(1):539–577, 2002.

E. Rozo, R. H. Wechsler, E. S. Rykoff, J. T. Annis, M. R. Becker, A. E. Evrard, J. A. Frieman, S. M. Hansen, J. Hao, D. E. Johnston, et al. Cosmological constraints from the sloan digital sky survey maxbcg cluster catalog. *The Astrophysical Journal*, 708(1):645, 2009.

B. Sartoris, A. Biviano, C. Fedeli, J. Bartlett, S. Borgani, M. Costanzi, C. Giocoli, L. Moscardini, J. Weller, B. Ascaso, et al. Next generation cosmology: constraints from the euclid galaxy cluster survey. *Monthly Notices of the Royal Astronomical Society*, 459(2):1764–1780, 2016.

R. Scaramella, J. Amiaux, Y. Mellier, C. Burigana, C. Carvalho, J.-C. Cuillandre, A. Da Silva, A. Derosa, J. Dinis, E. Maiorano, et al. Euclid preparation-i. the euclid wide survey. *Astronomy & Astrophysics*, 662:A112, 2022.

P. Schneider. Detection of (dark) matter concentrations via weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 283(3):837–853, 1996.

P. Schneider. *Extragalactic astronomy and cosmology: an introduction*, volume 146. Springer, 2006.

P. Schneider and C. Seitz. Steps towards nonlinear cluster inversion through gravitational distortions. i. basic considerations and circular clusters. *arXiv preprint astro-ph/9407032*, 1994.

V. Springel. The cosmological simulation code gadget-2. *Monthly notices of the royal astronomical society*, 364(4):1105–1134, 2005.

V. Springel, S. White, G. Tormen, and G. Kauffmann. Populating a cluster of galaxies-i. results at [formmu2] z= 0, mnras 328 (dec., 2001) 726–750. *arXiv preprint astro-ph/0012055*, 2001a.

V. Springel, N. Yoshida, and S. D. White. Gadget: a code for collisionless and gasdynamical cosmological simulations. *New Astronomy*, 6(2):79–117, 2001b.

R. Sunyaev and Y. B. Zeldovich. The spectrum of primordial radiation, its distortions and their significance. *Comments on Astrophysics and Space Physics, Vol. 2, p. 66*, 2:66, 1970.

J. L. Tinker, B. E. Robertson, A. V. Kravtsov, A. Klypin, M. S. Warren, G. Yepes, and S. Gottlöber. The large-scale bias of dark matter halos: numerical calibration and model tests. *The Astrophysical Journal*, 724(2):878, 2010.

K. Umetsu. Cluster–galaxy weak lensing. *The Astronomy and Astrophysics Review*, 28(1):1–106, 2020.

A. Vikhlinin, A. Kravtsov, R. Burenin, H. Ebeling, W. Forman, A. Hornstrup, C. Jones, S. Murray, D. Nagai, H. Quintana, et al. Chandra cluster cosmology project iii: cosmological parameter constraints. *The Astrophysical Journal*, 692 (2):1060, 2009.

A. M. Wolfe. *Perturbations of a cosmological model and angular variations of the microwave background*. The University of Texas at Austin, 1967.

C. O. Wright and T. G. Brainerd. Gravitational lensing by nfw halos. *The Astrophysical Journal*, 534(1):34, 2000.

Y. B. Zeldovich. A hypothesis, unifying the structure and the entropy of the universe. *Monthly Notices of the Royal Astronomical Society*, 160(1):1P–3P, 1972.

F. Zwicky. The redshift of extragalactic nebulae. *Helvetica Physica Acta, Vol. 6, p. 110-127*, 6:110–127, 1933.

F. Zwicky. Nebulae as gravitational lenses. *Physical Review*, 51(4):290, 1937.