



Estimation Of The Hubble Parameter Using Dark Sirens In Light Of New Generation Instruments

Scuola di Dottorato in Fisica
Dottorato di Ricerca in Fisica (XXXVI cycle)

Raul Ciancarella

ID number 775

Advisor
Enzo Franco Branchini

Co-Advisors
Carmelita Carbone
Marco Bonici

Academic Year 2022/2023

Thesis not yet defended

**Estimation Of The Hubble Parameter Using Dark Sirens In Light Of New
Generation Instruments**

PhD thesis. Roma Tre University of Rome

© 2024 Raul Ciancarella. All rights reserved

This thesis has been typeset by L^AT_EX and the Sapthesis class.

Author's email: raul.ciancarella@uniroma3.it



*Allegria nei giorni di sole.
Malinconia nei giorni di pioggia.
Se così ci viene insegnato, di questo ci convinciamo."*

—Neon Genesis Evangelion ep.26 La bestia che urlò Amore nel cuore del Mondo—

Abstract

In this thesis, I will explore the potential of cross correlating two distinct astronomical datasets: catalogues of galaxies acquired from spectroscopic redshift surveys and samples of gravitational wave (GW) events lacking electromagnetic counterparts. The aim is to assess the validity of the standard cosmological model and derive its fundamental parameters.

While the analysis of galaxies' spatial distribution and clustering properties has proven to be a powerful cosmological tool, evidenced by ongoing observational endeavours such as Boss, Euclid, DESI, Roman, and Rubin [1, 2, 3], the utilisation of GWs for cosmological inquiries is relatively new but increasingly intriguing since the inaugural direct detection of a gravitational signal by the LIGO collaboration [4].

There are two primary reasons for this increasing interest. Firstly, GWs offer a novel type of astrophysical messenger, wholly independent of photons, which have historically served as the sole informational conduit for astronomical investigations. Secondly, GWs can serve as standard "sirens" – extra galactic sources whose distance can be directly inferred from the observed signal.

Should the redshift of the GW host galaxy become observable, a classical Hubble test [5] could be conducted to deduce cosmological parameters, akin to the methodology employed with standard candles like Type Ia supernovae. Regrettably, the majority of detected GWs to date are "dark" – lacking unequivocal identification of a host galaxy and thus precluding redshift measurement. Nonetheless, the redshift of these dark sirens (DSs) can be statistically deduced by cross-correlating their spatial positions with those of galaxies in spectroscopic surveys, potentially serving as DS hosts, enabling a statistical rendition of the classical Hubble test.

Recent efforts have delved into this prospect, utilising hierarchical Bayesian statistical analyses to infer the Hubble constant's value [6, 7, 8]. These endeavours predominantly relied on existing data, comprised largely of local DS samples, due to the limited sensitivity of current GW detectors, and correlated with galaxy samples from disparate datasets.

The thesis aims to forecast future experiments' outcomes. A similar methodology has been recently employed in [9], who devised specialised software CHIMERA, to quantify the feasibility of measuring cosmological parameters using forthcoming O5 and O6 datasets from the LIGO-VIRGO-Kagra experiment.

This study will further expand upon this framework by contemplating the combination of the spectroscopic galaxy survey conducted by the Euclid satellite (scheduled for completion in 2029) with the catalogue of DSs anticipated from the planned Einstein Telescope experiment. Unlike prior investigations, this joint analysis will enable the execution of the Hubble test across a substantially broader volume, spanning the unprecedented redshift range $0.9 \leq z \leq 2$.

To this end, I have developed a proprietary data analysis pipeline, the Cross correlation with Dark sirEns (CODE). Since the absence of data for both channels, the first part of CODE is devoted to produce an observed Einstein Telescope-like catalogue of possible DSs, that will be cross correlated with the Euclid Flagship mock galaxy survey. The current version of CODE focus on deriving the Hubble

constant independently of cosmic microwave background (CMB) observations and local distance indicators. The reason for focusing on this measurement is the existing discrepancies in current H_0 measurements – the so-called Hubble tension – which constitutes a potential challenge to the foundation of the Λ -CDM model.

This projects methodically explores the effect that various factors, such as the accuracy of the distance measurements of Dark Sirens and the number of events, have on the final posterior, progressively increasing the level of complexity and realism of the problem. In the case of Euclid correlated with the Einstein Telescope, it is demonstrated how it is possible to achieve accuracy on the order of one percent, culminating in the determination of H_0 with an error of 0.5% after five years of data collection.

It's noteworthy that while this Thesis employs the pipeline to address a specific issue, its adaptable structure permits the handling of diverse datasets and the inference of various cosmological parameters, thereby facilitating exploration of alternative cosmological scenarios beyond the standard model.

Contents

Introduction	vi
1 Background Cosmology	5
1.1 Background Metric	5
1.1.1 FLRW metric	6
1.1.2 The Hubble constant	8
1.1.3 Background Evolution	10
1.1.4 Matter-Energy Content	13
1.2 The Hubble Tension	16
1.2.1 Hubble Constant from CMB	17
1.2.2 Hubble Constant from Local Distance Indicators	19
1.2.3 Two Headed Constant	21
2 Gravitational Waves	23
2.1 Deriving the Wave equation	23
2.2 Gauge Invariance and the Traceless Transverse Gauge	26
2.3 Solutions In The Low Velocity Regime	28
2.4 Momenta And Quadrupole	29
2.5 Binary Systems	31
2.6 Energy of GWs	34
2.7 Propagation	35
3 Poking Around The Dark	37
3.1 Gravitational Waves Detectors	37
3.1.1 Ground Based Interferometers and The Einstein Telescope	38
3.1.2 Antenna Pattern	41
3.2 Euclid	48
3.2.1 FlagShip	49
4 The Hierarchical Bayesian Formalism	53
4.1 A New Tool	54
4.1.1 Hierarchical inference	56
4.1.2 Galaxy Catalogue	58
4.1.3 The Beta normalisation	60
4.1.4 Analytical Insights	68

5	The CODE Pipeline	73
5.1	The Parent Catalogues	74
5.2	The Dark Sirens Catalogue	75
5.3	Making the Host Galaxy Catalogues	79
5.4	Estimating the Likelihood	80
5.5	Testing the Pipeline: Robustness tests.	82
5.5.1	Robustness to sample selection effects	82
5.5.2	Robustness to the choice of the galaxy host sample	84
5.5.3	Robustness to the choice of the DS sample	85
5.5.4	Robustness to the redshift distribution of the host galaxies	86
5.5.5	Robustness to the radial distribution of the DS	89
5.6	Forecasting	90
5.6.1	Sensitivity to the DS angular position uncertainties	91
5.6.2	Sensitivity to the uncertainty in the DS Luminosity Distance	92
5.6.3	Dependence on the number of dark sirens.	95
5.7	Forecast: 5 years of ET observation cross-matched with the Euclid spectroscopic galaxy catalogue	99
	Conclusions	114
	Acknowledgements	114
	Bibliography	115

Introduction

The Universe, its contents, and its laws have always fascinated humanity, representing the ultimate conundrum. Throughout human history, the firmament has served as both a map and a clock, with every civilisation looking to the stars for answers to its deepest questions. It is no wonder that, until relatively recent times, the study of the cosmos was linked to both philosophical and theological inquiries.

Modern cosmology aims to study the Universe as a whole, with a particular focus on its origin and evolution. The foundations of modern cosmology, as we know it, can be traced back to the first half of the last century thanks to the profound contributions of Albert Einstein and his theory of general relativity (GR) [10], as well as the observations of Lemaître and later Hubble [5].

First, in fact, a seemingly obvious observation with profound cosmological implications was performed earlier on by H. Olbers [11]. It goes under the name of *Olbers' paradox* and can be stated as follows: "Why is the night sky dark?" If we assume that the Universe is indeed infinite, then the surface brightness of the sky contributed by an infinite number of stars should be infinitely large. Its solution requires that the Universe is expanding, as observed by Hubble, so that the light of the distant sources could be redshifted away.

The observed expansion of the Universe with the implication that there is no preferential observer inspired the Cosmological Principle, i.e. that the Universe is an isotropic homogeneous physical system. Its assumption in the GR framework has led to the adoption of a Friedmann Lemaître Robertson Walker metrics [12] in the Einstein equations, whose solution constitutes the so-called Hot Big Bang model that we are still using.

This model has evolved into the current "standard" Λ CDM model to account for the intriguing, and largely unexplained, presence of two dark components: dark matter and dark energy, whose very nature constitute a challenge to fundamental physics.

The adoption of the Λ CDM model, despite the fundamental question it rises, is justified by its spectacular success in matching observations over a huge range of spatial and temporal scales.

Our recently achieved ability to reduce observational uncertainties to a per-cent level, allowed us to spot a few inconsistencies, usually referred to as "tensions" whose existence challenges the Λ CDM model. The most important one being the Hubble tension and involving the estimate of the current expansion rate of the Universe, quantified by the Hubble constant, obtained from different cosmological probes

[13, 14].

The quest for the value of the Hubble constant, H_0 , has a long history [15]. Its direct measurement in the local Universe can be obtained by comparing the redshift of an extra-galactic object to its distance. While redshift is readily obtained from the measured wavelengths of known lines in the observed energy spectrum of the object, estimating its distance is comparatively more challenging and can be only done for specific classes of objects called distance indicators. A distance indicator is an object for which a relation exists between two or more quantities (i.e. luminosity and rotation velocity for disk galaxies) one of which is related to a distance-dependent observable quantity i.e. luminosity and flux). The simplest example of distance indicators are the "standard candles". These are objects of known luminosity for which the measurement of the flux allows us to infer a particular type of distance, the luminosity distance, simply related to the proper distance of the object. The precision of these local measurements of H_0 has increased, reaching the per-cent level. The recent estimate [16] being:

$$H_0^{local} = (73.29 \pm 0.90) \text{ km s}^{-1} \text{ Mpc}^{-1} .$$

The second, completely different and somewhat model-dependent estimate of H_0 , can be obtained by exploiting the simple physics of the Universe at the last scattering epoch, when photons were last coupled to baryons. The statistical analysis of the temperature fluctuations in the Cosmic Microwave Background (CMB) allows one to extract a wealth of cosmology-relevant information from which one can infer the H_0 value. The estimate obtained from the analysis of the CMB maps produced by the Planck satellite mission is also very precise, but significantly different from the local measurement [17]

$$H_0^{planck} = (67.66 \pm 0.62) \text{ km s}^{-1} \text{ Mpc}^{-1} ,$$

The discrepancy, the Hubble tension, is currently considered a significant issue in cosmology and poses a potential threat to the Λ CDM model. Despite efforts to identify possible sources of systematic errors in the data and analyses, several alternative models have been proposed to alleviate this tension [13]. None of these models have been able to fully explain the observed discrepancy.

A game-changer in this state of affairs would be represented by an independent method of estimation derived from a completely different cosmological probe. This possibility, first contemplated by B. Schutz [18], eventually materialised in 2015 with the detection of the first gravitational wave by the LIGO observatory [19]. The existence of gravitational waves (GWs) was theorised by Einstein in the early 20th century [20] These ripples in the space-time fabric are produced, for example, when two massive objects whirl toward each others.

The simultaneous observation of such an event, triggered by the coalescence of two neutron stars, and its electromagnetic counterpart confirmed that GWs travel at the speed of light. Once detected, the GW signal carries crucial information

on the intrinsic magnitude of the event. When compared with the intensity of the signal, it allows us to infer the distance of the gravitational waves. For these reasons, gravitational waves constitute a new class of distant indicators, dubbed "standard sirens."

Like all other distance indicators, to measure H_0 one needs to estimate the distance and measure the redshift of the GW source. However, this has been possible in only one case [4, 21], out of the ~ 90 GW events detected so far by the LIGO Virgo Team [22].

The paucity of observed GWs with electromagnetic counterpart, also known as "bright sirens", does not allow estimating H_0 with the required precision, nor the current rate of bright sirens detection allows us to predict when this will be possible.

A viable alternative to bright sirens (BS) are the dark sirens (DS), i.e. GWs without electromagnetic counterpart. These represent the vast majority of detected events, but with no observed redshift available, do not allow us to estimate H_0 directly. An indirect estimate is, however, possible by statistically correlating the observed position of the DSs with that of the surrounding potential hosts, i.e. with the position of luminous galaxies in large spectroscopic surveys.

Redshift surveys have revolutionised our understanding of the Universe, offering a detailed mapping of galaxy distribution and large-scale structure. Pioneering studies such as the Sloan Digital Sky Survey [23] (SDSS), the Baryon Oscillation Spectroscopic Survey [24] (BOSS), and its successor, the extended Baryon Oscillation Spectroscopic Survey [25] (eBOSS), have provided unprecedented insights into the Universe's structure. BOSS, in particular, has enabled measurements of the baryon acoustic oscillations [26] (BAO) scale with unparalleled precision, becoming a cornerstone in studying dark energy and dark matter in the Universe. Its successor, eBOSS, has extended this mapping to higher redshifts, thus offering new opportunities to explore the Universe's evolution.

Ongoing surveys like the Dark Energy Spectroscopic Instrument [2] (DESI) and Euclid are pushing the boundaries of knowledge even further. DESI, with its ability to measure about 30 million galaxy and quasar redshifts, aims to construct a three-dimensional map of the Universe with unprecedented precision. However, it is the European Space Agency's Euclid project that promises to be revolutionary [1, 27]. Designed to survey the sky with unmatched precision and depth, Euclid aims to unveil the secrets of dark energy and the Universe's structure. Through its observations, Euclid will contribute to answering fundamental questions about the nature of cosmic expansion and the formation of cosmic structures. Euclid's importance in cosmology is indisputable, and its launch is eagerly anticipated by the scientific community for its potential discoveries and contributions to our understanding of the Universe.

The significance of GWs extends well beyond the sole determination of H_0 . GWs can also be used to determine other cosmological parameters that appear in the $z - D_l$ relationship, in addition to providing a new perspective from which to study alternative theories of gravity [6]. Cosmologically, they can serve as a means to identify and study primordial black holes [28], while from an astrophysical standpoint, GWs are an invaluable source of information about the internal structure of the

source, allowing for the detection of possible exotic compact object. [29, 30, 31, 32].

The goal of this thesis is to assess the accuracy with which cosmological parameters, with an emphasis on H_0 , can be measured by cross-correlating catalogues of dark sirens generated by current and future gravitational wave detectors with those of galaxies obtained from ongoing observational spectroscopic surveys. For this we have set up a data analysis pipeline consisting of a few key elements. First, we generate mock catalogues of dark sirens and galaxies mimicking the datasets that will be produced by the next generation GW interferometers like the Einstein Telescope [33] (ET) and by the Euclid galaxy survey [27]. We implement and perform a Bayesian inference analysis similar to the one adopted by [6] to estimate the posterior probability of the cosmological parameters of interest given by the two datasets. While in this thesis, I will focus on the ET and Euclid samples on one hand and on the Hubble constant on the other hand, the pipeline is flexible enough to provide forecasts for additional cosmological parameters using other datasets, including existing ones.

The layout of the thesis is as follows: The first chapter will introduce the necessary notions of cosmology and will delve into the Hubble tension, current measurements, and methods found in the literature.

The second chapter is dedicated to GWs (Gravitational Waves): from the mathematical derivation to their propagation in an FRW metric. The quadrupolar nature of GWs will be explained, and a binary system will be briefly studied, with an emphasis on the frequency and energy of GWs.

The third chapter, on the other hand, presents the "tools" on which the analysis has focused. The first part discusses the ET and Euclid experiments, with particular attention to the former, as it relates to the properties of the sources that are actually visible.

In the fourth chapter, the theoretical framework necessary to conduct the study is discussed. The concept of hierarchical Bayesian inference is briefly introduced, and relevant physical quantities are discussed, with a final digression on the analytical properties of some functions.

The fifth chapter encompasses the operational scheme of the pipeline, robustness tests and calibration, along with results for five years of observation assuming ET-like and Euclid-like instruments.

In the conclusion, a brief summary of the work can be found.

Chapter 1

Background Cosmology

In this chapter, I will introduce the scientific concepts and tools that constitute the building blocks of this work. Cosmology aims to study the Universe as a whole, i.e. out to the largest observable scales¹ The Universe operates as a self-gravitating system, where, on sufficiently large scales, all other forces can be neglected. Consequently, comprehending its evolution necessitates a theory of gravity. In this context, Einstein's General Relativity is adopted, and the goal is to solve Einstein's equation.

Solving these coupled, nonlinear partial differential equations is a challenging endeavour that necessitates adopting some approximation for space-time geometry. Observations provide evidence that the Universe is spatially homogeneous and isotropic on sufficiently large scales. We elevate this evidence to a Cosmological Principle and construct a model for the background, upon which cosmic structures form and evolve.

1.1 Background Metric

As anticipated, the model for background cosmology rests upon the Theory of General Relativity, formulated by Einstein in 1915. This theory is built upon three pillars:

- **The Equivalence Principle:** This principle asserts that the effect of gravity is indistinguishable from that of an acceleration or, alternatively, that no experiments can distinguish between a uniformly accelerated system and a system in a gravity field.
- **Covariance:** The theory of General Relativity adheres to the idea of covariance, ensuring that all observers will obtain identical results for a given physical phenomenon. This consistency extends across reference frames, necessitating appropriate adjustments to equations based on the chosen frame of reference.

¹Here, and in the rest of the work, the term "large scale" indicates a scale of ~ 100 Mpc that safely encompasses the size of the largest cosmic structures.

- **Four-Dimensional space-time:** General Relativity applies to a four-dimensional space-time, encompassing three spatial dimensions and time. The geometry of the system, particularly its curvature, is connected to its energy content through Einstein's equations.

A crucial quantity for capturing the properties of space-time is the metric tensor, since it allows us to define distances in a curved space-time. In a Euclidean, flat space the metrics, defined as the separation between two nearby points is

$$ds^2 = dx^2 + dy^2 + dz^2 = \sum_{i,j=0}^2 g_{ij} dx^i dx^j, \quad (1.1)$$

where ds is the spatial separation and g_{ij} is defined as

$$g_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \quad (1.2)$$

that is a Kronecker delta δ_{ij} . The explicit expression of the metric tensor depends on the coordinate system. In the spherical coordinate system the metric element is

$$ds^2 = dr^2 + r^2(d\theta^2 + \sin^2 \theta d\varphi^2). \quad (1.3)$$

The corresponding metric tensor $g_{ij} = \text{diag}(1, r^2, r^2 \sin^2 \theta)$ is still diagonal, but its coefficients depend on the coordinate, even in absence of curvature, since ds^2 is invariant.

Special Relativity describes physical systems in a 4-dimensional spacetime with Minkowski metrics

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 = \sum_{i,j=0}^3 \eta_{ij} dx^i dx^j = \eta_{ij} dx^i dx^j, \quad (1.4)$$

where the symbol η identifies the Minkowski metric tensor of the flat 4-dimensional spacetime. Self gravitating systems, like the Universe, can be described by the Minkowski metrics only locally. A global description requires using a different metrics, even in absence of curvature.

1.1.1 FLRW metric

The adoption of the Cosmological Principle implies assuming several space-time symmetries that greatly simplify its metrics. Homogeneity and isotropy imply that the metric tensor is diagonal, and the $_{00}$ coefficient of the time element dt does not depend on the spatial coordinate; thus, it can be factorised out:

$$ds^2 = -dt^2 + a^2(t) dl^2, \quad (1.5)$$

where dl^2 is the line element for the spatial part of the metric and $a(t)$ is called *scale factor*. To preserve homogeneity and isotropy, the scale factor must be a function of

time only. Its physical meaning is clear: the relative spatial distances between any two points increase (or decrease) as $a(t)$, implying a global expansion (contraction) of the entire system. Thus, the Hubble expansion is an intrinsic feature of any homogeneous, isotropic, self-gravitating system.

We still have to specify the spatial line element dl . Curvature in a generic space-time is a complex quantity fully captured by the rank-4 Riemann tensor. In the highly symmetric system we are considering here, curvature is described by a time-independent scalar, constant throughout the spacetime. Its sign fully determines the geometrical property of the system:

- **Positive Curvature and Spherical Geometry:** In this case, constant time hypersurfaces are represented by a three-sphere in a Euclidean E^4 space. The curvature is the radius R of the sphere in this four dimensional space

$$dl^2 = du^2 + d\mathbf{x}^2, \quad u^2 + \mathbf{x}^2 = R^2.$$

- **Zero Curvature and Flat Geometry:** In this case, the line element is a three-dimensional Kronecker delta $dl^2 = \delta_{ij}dx^i dx^j$ describing the properties of the familiar Euclidean geometry.
- **Negative Curvature and Hyperbolic Geometry:** In this case constant time hypersurfaces are three-dimensional hyperboloid embedded in what is called a Lorentzian four dimensional space, meaning that one coordinate has to have an opposite sign

$$dl^2 = -du^2 + d\mathbf{x}^2, \quad -u^2 + \mathbf{x}^2 = -R^2, \quad R^2 > 0.$$

In polar coordinate one can express the metric element as

$$ds^2 = -dt^2 + a^2(t) \left(\frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right), \quad (1.6)$$

where $d\Omega^2 = d\theta^2 + \sin^2(\theta)d\phi^2$ and k is the curvature that, after conveniently rescaling the coordinate, can have the following values:

$$\begin{cases} k = +1 & \textit{Spherical} \\ k = 0 & \textit{Flat} \\ k = -1 & \textit{Hyperbolic} \end{cases} . \quad (1.7)$$

(r, θ, ϕ) , or equivalently (x, y, z) in a Cartesian system, are called co-moving coordinates and are dimensionless. The expansion factor $a(t)$ has the dimension of a length. Its value can be conveniently normalised to the present epoch, i.e. we can set $a(t = 0) = 1$.

Equation 1.6 provides the expression of the Friedmann Lemaître Robertson Walker (FLRW) metrics. It is derived from the Cosmological Principle and assumed in the standard cosmological model considered in this Thesis.

1.1.2 The Hubble constant

As anticipated, the FLRW metrics implies that the Universe is expanding, as observed by Hubble. To see this more explicitly and provide a quantitative definition of the Hubble constant, let us consider two nearby points in spacetime on a constant time hypersurface $dt = 0$. Because of isotropy one can consider their separation along the radial direction $d\theta = d\phi = 0$ and find

$$dl(t) = a(t) \left(\frac{dr^2}{1 - kr^2} \right)^{1/2} \equiv a(t)F(k, r)dr. \quad (1.8)$$

Since the separation changes over time a relative velocity of the two points can be defined by deriving dl with respect to the time coordinate

$$v \equiv \dot{dl} = \dot{a}F(k, r) = \frac{\dot{a}(t)}{a(t)}aF(k, r) \equiv H(t)dl$$

which is valid for any two points at any finite separation. It describes the recession velocity of galaxies that, as observed by Hubble, increases linearly with their distance. The time-function $H(t)$ is the so-called Hubble parameter, and its value at the present epoch H_0 is the quantity I will focus on in this thesis. This value, which quantifies the rate at which the Universe is expanding, or contracting, is expressed in units of $\text{Km}, \text{s}^{-1}, \text{Mpc}^{-1}$.

While the recession velocity can exceed the speed of light, it does not violate causality. For any observer, the distance at which the recession velocity equals the speed of light represents its Hubble horizon, $\lambda_H = H^{-1}$ in natural units.

Cosmological Redshift

One consequence of the FLRW metrics is cosmological redshift. The energy of a photon travelling between a source and an observer decreases (or increases) depending on their relative distance or, since photons travel at the speed of light in empty space, on the time delay between the emission of the photon and its detection. Due to this characteristic effect, redshift is used as both a time proxy (indicating that a source at high redshift emitted the photon long ago) and a distance proxy (indicating that a source at high redshift is far away).

It is possible to show that the energy E of a photon travelling in an FLRW spacetime changes with time as

$$\frac{1}{E} \frac{dE}{dt} = -\frac{\dot{a}}{a}. \quad (1.9)$$

Since for a photon $E \propto \lambda^{-1}$ the above relation can be rewritten as

$$\frac{d\lambda}{\lambda} = -\frac{da}{a}. \quad (1.10)$$

which is valid for any finite separation and time interval, leading to

$$z \equiv \frac{\lambda_0 - \lambda_e}{\lambda_e} = \frac{1}{a(t_e)} - 1, \quad (1.11)$$

where the subscripts $_0$ and $_e$ indicate the observation and emission time, respectively, and the normalisation $a(t = 0) = 1$ has been used. The quantity z is the cosmological redshift. It can be estimated by observing the wavelength of photons whose energies at the emission epoch are known. This can be done by identifying characteristic emission or absorption lines in the energy spectrum of extra galactic objects. The inferred quantity is dubbed 'spectroscopic redshift.' Alternatively, a 'photometric redshift' can be observed by measuring the energy flux in various bands, effectively coarsely sampling the energy spectrum of the source. Spectroscopic redshifts provide a more precise estimate of the cosmological redshift than photometric ones.

It is the cosmological redshift, sometimes incorrectly interpreted as Doppler shift, that has been observed by Hubble to measure the galaxies' recession velocity and discover its eponymous relation.

Distances in Cosmology

Having defined a metric for spacetime, we can define distances between any two points. The proper distance between them is defined over a constant time hypersurface, i.e., having imposed $dt = 0$. It can be estimated, thanks to isotropy and homogeneity, along a radial path connecting the two points by placing one of them at the centre of coordinates. In comoving units, its expression is:

$$d_c = \int_r^0 a(t) F(k, r) dr. \quad (1.12)$$

Its value depends on the curvature of the system. While well-defined, the proper distance is not a measurable quantity since it would require performing measurements instantaneously ($dt = 0$). However, d_c is also equal to the comoving distance travelled by a photon emitted at time t_e from r and received by the observer at 0 at t_0 :

$$d_c = \int_{t_e}^{t_0} \frac{1}{a(t)} dt = \int_0^{z_e} \frac{1}{H(z)} dz, \quad (1.13)$$

where we have exploited the fact that photons travel along null paths ($ds = 0$) and have used eq. (1.11) in the last expression eq. (1.12) to change the variable from t to z .

This second expression brings us closer to a measurable quantity, the *luminosity distance*. Let us consider the energy flux F collected by an observer at t_0 and emitted by an extra-galactic source with an intrinsic luminosity L . The flux decreases with the square of the luminosity distance of the source:

$$F = \frac{L}{4\pi D_l^2}, \quad (1.14)$$

which is simply related to the comoving distance of the source $D_l = (1+z)d_c$. Thus, the possibility of measuring D_l and the redshift z allows us to infer the comoving distance of the source d_c and, through eq. (1.13), the Hubble function $H(t)$.

To measure the luminosity distance, we therefore need to measure the flux of a source whose intrinsic luminosity is known. This cannot be done for generic sources but only for distance indicators. One historically relevant example, particularly for the local measurement of H_0 , is the class of Cepheid stars. Cepheids are bright, variable stars that pulsate radially, varying both their diameter and temperature, and consequently, their brightness. The observed light curve has a well-defined stable period that is a function of the intrinsic luminosity [34, 35]. Cepheids can be detected and observed within relatively nearby galaxies, reaching out to approximately 50 Mpc.

A second type of much brighter distance indicators, historically relevant for the measurement of H_0 and, as we shall see, for probing the accelerated expansion of the Universe, is Type Ia Supernovae.

Type Ia supernovae are a particular type of supernovae that occur in a binary system composed of a normal star and a white dwarf. The white dwarf accretes mass from its companion, and once the critical mass, that is the Chandrasekhar limit, is reached, nuclear fusion sets in an uncontrolled explosive way, resulting in a disruption of the system.

Since the Chandrasekhar limit is a fixed mass of $\sim 1.4, M_\odot$ ², the intrinsic peak luminosity must be similar for all Type Ia Supernovae, irrespective of the characteristics of the original binary system. In fact, Type Ia supernovae do not all reach the same peak luminosity. However, the width of the observed light curve can be used to correct Type Ia supernovae to standard candle values. The original correction, known as the Phillips relationship, makes Type Ia supernovae very precise standard candles with which it is possible to measure relative distances to $\sim 10\%$ accuracy [36, 36].

1.1.3 Background Evolution

Now that the geometry of the isotropic homogeneous Universe is defined, it is time to link its geometric properties to the mass-energy content and derive the equations that describe its dynamics.

The starting point are the fundamental Einstein's Field equations

$$G_{\mu\nu} = 8\pi T_{\mu\nu}, \quad (1.15)$$

where $T_{\mu\nu}$ is the stress-energy tensor, representing the contribution of all forms of energy and matter in the Universe, and $G_{\mu\nu}$ is the Einstein tensor, obtained by combining the metric tensor, the Ricci tensor and its contraction, the Ricci scalar

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R. \quad (1.16)$$

²there can be variation depending on the model, but the point is that the limit is the same for every white dwarf

The Ricci tensor contains information on the metric tensor and its first and second derivatives. It is defined as

$$R_{\mu\nu} = \frac{\partial \Gamma^\alpha_{\mu\nu}}{\partial x^\alpha} - \frac{\partial \Gamma^\alpha_{\mu\alpha}}{\partial x^\nu} - \Gamma^\alpha_{\sigma\alpha} \Gamma^\sigma_{\mu\nu} + \Gamma^\alpha_{\sigma\nu} \Gamma^\sigma_{\mu\alpha}, \quad (1.17)$$

where $\Gamma^\alpha_{\mu\nu}$ are the Christoffel symbols:

$$\Gamma^\alpha_{\mu\nu} = \frac{1}{2} g^{\alpha\gamma} \left(\frac{\partial g_{\gamma\nu}}{\partial x^\mu} + \frac{\partial g_{\mu\gamma}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\gamma} \right). \quad (1.18)$$

Assuming isotropy and homogeneity and under the hypothesis that all forms of matter and energy relevant to cosmology can be described as a perfect fluid, the stress-energy $T_{\mu\nu}$ is obtained, and can be cast in the following form:

$$T_{\mu\nu} = (\rho + P)U_\mu U_\nu + P g_{\mu\nu}, \quad (1.19)$$

where ρ indicates the energy density in the fluid rest frame, P its isotropic pressure, and U_α is the four-velocity of the fluid element with respect to the observer. The stress-energy tensor must also obey the conservation equation:

$$0 = \nabla_\mu T^\mu_\nu = \frac{\partial T^\mu_\nu}{\partial x^\mu} + \Gamma^\mu_{\alpha\mu} T^\alpha_\nu - \Gamma^\mu_{\nu\mu} T^\mu_\alpha. \quad (1.20)$$

These are four differential equations, and the first one, $\nu = 0$, reads

$$0 = -\frac{d\rho}{dt} - \Gamma^\mu_{\mu 0} \rho - \Gamma^\alpha_{\mu 0} T^\mu_\alpha. \quad (1.21)$$

Its explicit expression for the LFRW metrics is

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + P) = 0, \quad (1.22)$$

which describes the usual continuity equation for a perfect fluid in an expanding, or contracting, system.

The two quantities ρ and P are related through the equation of state of a perfect fluid. The equation of state for cosmologically relevant fluids is usually parameterised as $P = \omega\rho$. The parameter ω is dubbed the 'Zel'dovich parameter,' and its value fully characterises the physical properties of the fluid. Inserting the equation of state into equation 1.22, the following is obtained:

$$\frac{\dot{\rho}}{\rho} = 3(1 + \omega)\frac{\dot{a}}{a}. \quad (1.23)$$

Its solution

$$\rho \propto a^{3(1+\omega)} \quad (1.24)$$

indicates that the energy density of the fluid decreases as the Universe expands, unless $\omega \leq -1$. The rate at which it decreases depends on the characteristics of the fluid. For non-relativistic matter, describable as a pressureless fluid with $\omega = 0$,

the mass density decreases as a^{-3} , meaning it is diluted by the increase in the physical volume of the fluid element. The energy density of radiation, characterised by $\omega = 1/3$, decreases as a^{-4} . The additional a^{-1} dilution factor quantifies the effect of cosmological redshift.

Finally, observational evidence of an accelerated expansion of the Universe requires that the energy budget of the Universe is currently dominated by a new component that behaves as a perfect fluid with $\omega \leq -1/3$. The value $\omega \leq -1$, which provides a good fit to observational data, corresponds to the case of a cosmological constant and implies that the corresponding energy density does not evolve with time.

Different fluids, which are assumed to interact only through gravity, may contribute to the stress-energy tensor, each characterised by an equation of state 1.24. In this case, the pressure and energy density terms that appear in the tensor represent the sum of all these components.

The Friedmann Equations

While ρ and P are related through the equation of state, an additional equation is needed to complement 1.22 in order to solve for $a(t)$. This equation, known as the first Friedmann equation, is obtained by considering the $_{00}$ component of the Einstein equations:

$$\dot{a}^2 = H^2(t) = \frac{8\pi G}{3}\rho - \frac{k}{a^2}, \quad (1.25)$$

where, once again, the normalisation $a(t_0) = 1$ is used, and ρ accounts for the energy density of all fluids in the Universe. To identify them, the subscript $_r$ is used for radiation, $_m$ for non-relativistic matter, and $_\Lambda$ for the cosmological constant."

If the total energy density matched the critical one

$$\rho_{crit} \equiv \frac{3H}{8\pi G}. \quad (1.26)$$

then, according to equation 1.25, the curvature of the system must be null. It is convenient to express the energy density of each fluid in units of the critical density and define the density parameter:

$$\Omega_x = \frac{\rho_x}{\rho_{crit}}; \quad \text{where } x = r, m, \Lambda. \quad (1.27)$$

The addition of the subscript $_0$ indicates the quantity measured at the present epoch $t = t_0$. A density parameter for the curvature term can also be formally defined:

$$\Omega_k \equiv -\frac{k}{H^2}, \quad (1.28)$$

so that the following relation holds true:

$$\Omega_r + \Omega_m + \Omega_k + \Omega_\Lambda = 1. \quad (1.29)$$

This relation quantifies the fractional contribution of each fluid to the overall energy density of the Universe at any epoch.

The "second Friedmann equation" is obtained by any of the three (identical) ii spatial components of the Einstein's Equations

$$\frac{\ddot{a}}{a} = \frac{4\pi G}{3} (\rho + 3P) . \quad (1.30)$$

The two Friedmann equations and the energy conservation equation are not independent. Any two of them, along with the equation of states of all the fluids that contribute to the energy density of the system, can be solved for $a(t)$ and, consequently, model the expansion history of the Universe $H(t)$.

1.1.4 Matter-Energy Content

The expansion history of the Universe depends on the relative density and equation of states of the various fluids in the Universe. Such dependence shapes the relation between the measured luminosity distance and the redshift of the source z_s . This relationship becomes clear after plugging the first Friedmann equation into eq. (1.13), taking into account the relation between comoving and luminosity distances:

$$D_l = \frac{(1+z)_s}{H_0} \int_0^{z_{0s}} \frac{dz}{\sqrt{\Omega_{r,0}(1+z)^4 + \Omega_{m,0}(1+z)^3 + \Omega_{k,0}(1+z)^2 + \Omega_{\Lambda,0}}} . \quad (1.31)$$

This relation stands as a pivotal aspect of my thesis, highlighting the capability to derive the value of the Hubble constant at the present epoch, H_0 , through the measurement of luminosity distance and redshift of a distance indicator. This inference is achievable when considering a cosmological model, completely defined by the density parameters $\Omega_{r,0}, \Omega_{m,0}, \Omega_{\Lambda,0}$, whose values are measurable through additional observations.

In the following, I shall review the individual contributions of the various fluids to the total energy density within the framework of the flat Λ CDM model, which allows us to set $\Omega_{k,0} = 0$.

Relativistic Particles

Stable, relativistic particles contribute to the energy density of the Universe. Thanks to the rapid decrease of their energy density, $\rho_r \propto a^{-4}$, relativistic particles dominate the energy budget of the Universe at early times. Cosmologically relevant relativistic particles include the photons of the cosmic microwave background and neutrinos before they become non-relativistic. The former exhibit an almost perfect black body spectrum with a temperature $T_{0,\gamma} = 2.72548 \pm 0.00057$ K [37], corresponding to a present day mean energy density $\Omega_{r,0} = (5.401 \pm 0.012) \times 10^{-5}$ [17].

Neutrinos are massive but light. They decouple before the electron-positron annihilation when they are still relativistic. Their temperature is smaller than that of the photons by a factor $\left(\frac{4}{11}\right)^{1/3}$. Their energy density is reduced accordingly and,

for this reason, they are always sub-dominant with respect to photons.

Non-relativistic matter

Non-relativistic particles behave as a pressureless fluid on cosmological scales, i.e., their equation of state is characterised by a Zel'dovich parameter $\omega = 0$, and therefore, their mass density decreases as a^{-3} with the expansion of the Universe.

Non-relativistic particles include standard *baryonic* matter, made of known particles. But these are not the only components. Another type of matter, *dark matter*, which manifests itself only through gravitational interaction so far, is assumed to be present and dominating the total mass budget.

The baryonic mass content has been precisely inferred from the abundance of light elements produced during cosmological nucleosynthesis ($\Omega_b h^2 = 0.020 \pm 0.002$, where h indicates the H_0 value in units of $100 \text{ km s}^{-1} \text{ Mpc}$; [38] and from the analysis of the brightness fluctuations of the CMB ($\Omega_b h^2 = 0.0224 \pm 0.0001$ [17]).

Local measurements of the baryon density have returned systematically lower values [39, 40], triggering the quest for the *missing baryons* that has indicated that a significant fraction of the local baryons are to be found in the form of a warm-hot intergalactic medium [41]. The first evidence that more matter, in addition to baryons, is present in the Universe traces back to 1933, in Zwicky's observations of the excess random velocity of galaxies within the Coma cluster [42]. However, the adoption of the Dark Matter hypothesis, i.e., that this excess was due to a new type of matter, was proposed in the '70s after the observations of the flat rotation curves of disk galaxies by V. Rubin [43]. An overwhelming body of evidence shows that dark matter exists and is made of massive, non-relativistic particles with negligible cross-section for interaction with standard particles, including photons.

Cold (i.e. non-relativistic at decoupling) Dark Matter is now firmly part of the standard cosmological model [44, 45] and the quest for its direct detection represents a very active research field.

With no direct detection so far, the nature of dark matter particles remains unknown. A plethora of dark matter candidates have been proposed, ranging from weakly interacting massive particles proposed in the framework of the supersymmetrical extensions of the standard model of particle physics [46] to the more recent proposal of primordial black holes [47].

The energy density of matter, both baryonic and dark, from [17] is $\Omega_m = 0.3111 \pm 0.0056$.

Cosmological Constant and Dark Energy

The discovery of the expanding Universe is one of the most significant achievements in modern cosmology, and it has been largely attributed to the observations of Type Ia Supernovae. These stellar phenomena have played a crucial role in revealing that the Universe is not static but is, in fact, expanding at an accelerating rate.

The breakthrough came in the late 1990s when two independent research teams, the Supernova Cosmology Project [48] and the High-Z Supernova Search Team [49],

used distant Type Ia Supernovae to measure the rate of expansion of the Universe. They observed that distant supernovae appeared fainter than expected under the presumption of a decelerating Universe, which would be the case if gravity was the sole force acting on large scales.

These observations imply that the Universe is accelerating its expansion, a behaviour utterly unexpected for a self-gravitating system. In Einstein's General Relativity framework, an accelerated expansion can only be accounted for if, at the present epoch, the energy budget of the Universe is dominated by a fluid with negative pressure and an equation of state with $\omega < -1/3$ [50], as shown by the second Friedmann's equation. This mysterious fluid is dubbed Dark Energy.

Among the many Dark Energy models that have been proposed so far, the simplest one that is consistent with all observations is the Cosmological Constant Λ .

The cosmological constant was first introduced by Einstein to make the Universe static and match the observational constraints at that time. After being abandoned following Hubble's discovery that the Universe is expanding, it has now been reconsidered to account for the accelerated expansion. The introduction of this constant modifies the field equations as follows:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}, \quad (1.32)$$

or, equivalently, as

$$G_{\mu\nu} = 8\pi G (T_{\mu\nu} + T_{\mu\nu}^{\Lambda}), \quad (1.33)$$

where we have defined $T_{\mu\nu}^{\Lambda}$ as

$$T_{\mu\nu}^{\Lambda} \equiv -\frac{\Lambda}{8\pi G} g_{\mu\nu}. \quad (1.34)$$

Having written Einstein's equations in this form, it is clear that the cosmological constant can be regarded as a fluid with its pressure and energy density that contributes to the total stress-energy tensor of the system. The equation of state of this fluid is characterised by a Zel'dovich parameter $\omega = -1$, which implies the remarkable property that its energy density does not evolve with time. Other forms of Dark Energy, characterised by different, and potentially time-dependent, values of ω , do not share this characteristic. Its energy density from [17] is $\Omega_{\Lambda} = 0.6889 \pm 0.0056$.

Total Mass Energy Budget

Putting all together, the total energy-mass content of the Universe at the present epoch can be determined. The results are summarised by the pie diagram in Fig. 1.1, where we show the per-cent contribution of the various components to the total energy budget, keeping in mind that in the Λ CMD framework adopted here the Universe is flat and, therefore, the total energy density must match the critical value.

Dark Energy dominates the budget. However, the relative contribution of the various components to the total energy budget evolves with time. Going back in time, the contribution of the relativistic species, negligible at the present epoch, increases and becomes the dominant one during the radiation epoch. In between

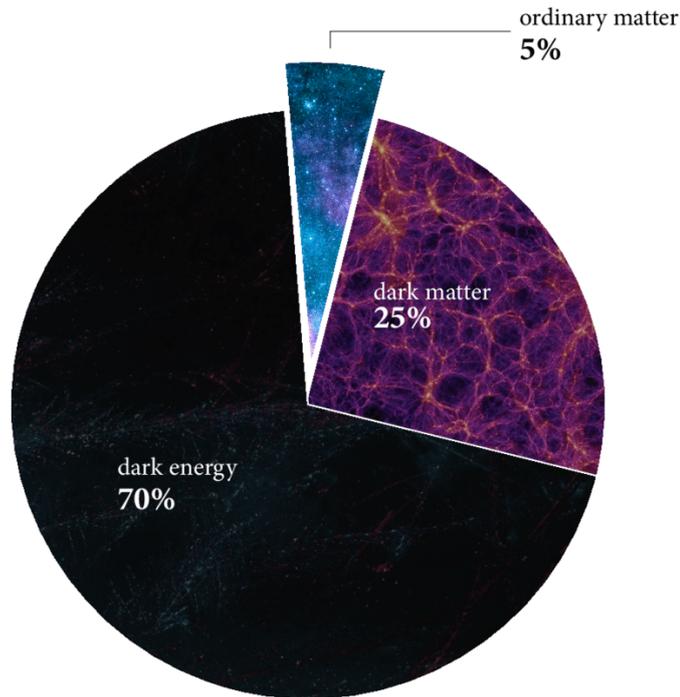


Figure 1.1. The image is a rough, but immediate representation of how much each component contributes to the overall energy budget. Source [51].

the radiative era and the dark energy-dominated one, the budget is dominated by the non-relativistic component, during the so-called matter era.

1.2 The Hubble Tension

Despite lacking an understanding of the nature of its dark components, the Λ CDM model successfully fits almost all the observational data. With a few exceptions, the most significant of which is the so-called Hubble tension.

Cosmological observations have revealed a discrepancy between the value of H_0 inferred from different types of observations. A discrepancy whose statistical significance has increased with the increasing accuracy of the observations themselves.

Being a factor that sets the length and timescale of the whole system, H_0 can be inferred from various types of observations. Two of them, the estimate of H_0 from X-ray observations and the Sunayev-Zel'dovich effect of galaxy clusters [52], and from strong lensing time delay [53], have not yet reached the required precision to play a relevant role in the current tension. The tensions stem from two entirely different types of observations: the analysis of the cosmic microwave background [17] and the measured distances of various distance indicators in the local Universe [54]. A detailed review on this topic can be found in [13, 55]. Here I simply present an overview of the problem.

1.2.1 Hubble Constant from CMB

Originating at the last scattering surface, approximately 380,000 years after the Big Bang, the CMB encodes pivotal information about the physics of the young universe. Recombination, a transformative event during this epoch, saw the formation of neutral hydrogen atoms as protons and electrons, previously in the form of a highly ionised plasma. After recombination, the Universe became transparent to radiation, allowing photons to traverse the cosmos unimpeded. As a result, CMB is the furthest observable phenomenon visible in the electromagnetic channel.

While incredibly uniform at large scale, the Universe presents some non-homogeneity that, thanks to gravitational instability, gave birth to the structures that can be seen today.

Those small deviations were most probably caused by quantum fluctuation of the primordial energy density distribution [56] and those primordial spikes are believed to be the seeds necessary to structure formation. Those seeds are over dense regions that grew thanks to gravity positive feedback. The field of structure formation is a complex and wide area of Cosmology, and I will not delve into the evolution of those scalar perturbations. The main interest of this work are tensorial perturbations i.e. the propagation of GWs, which will be discussed in the next chapter.

Scalar perturbation plays an important role in the quest for the H_0 parameter, since they encode precious information about the "background" Universe. For the scope of this work, the relevant aspect of those perturbations are the temperature fluctuation in the CMB spectrum. CMB photons carry all the physical information of the Universe at the last scattering epoch. Their observations by the COBE satellite [57] have revealed that this background is not completely isotropic. Instead, it possesses small temperature fluctuations $\delta T/T \sim \mathcal{O}(10^{-5})$, which constitute the seeds from which cosmic structures will arise later [58]. The sky map of these fluctuations, as observed by the Planck satellite mission [17], is shown in Fig. 1.2.

Those temperature fluctuations obey an almost Gaussian distribution, therefore they can be fully described by their *angular power spectrum*. The angular power spectrum $\mathcal{C}(l)$, projected on the celestial vault, can be decomposed using a multipolar expansion in spherical harmonics:

$$\mathcal{C}(l) = \frac{1}{2l+1} \sum_{m=-l}^l |Y_l^m|^2, \quad (1.35)$$

where $\mathcal{C}(l)$ is the angular power spectrum and Y_l^m is the spherical harmonic mode of order l .

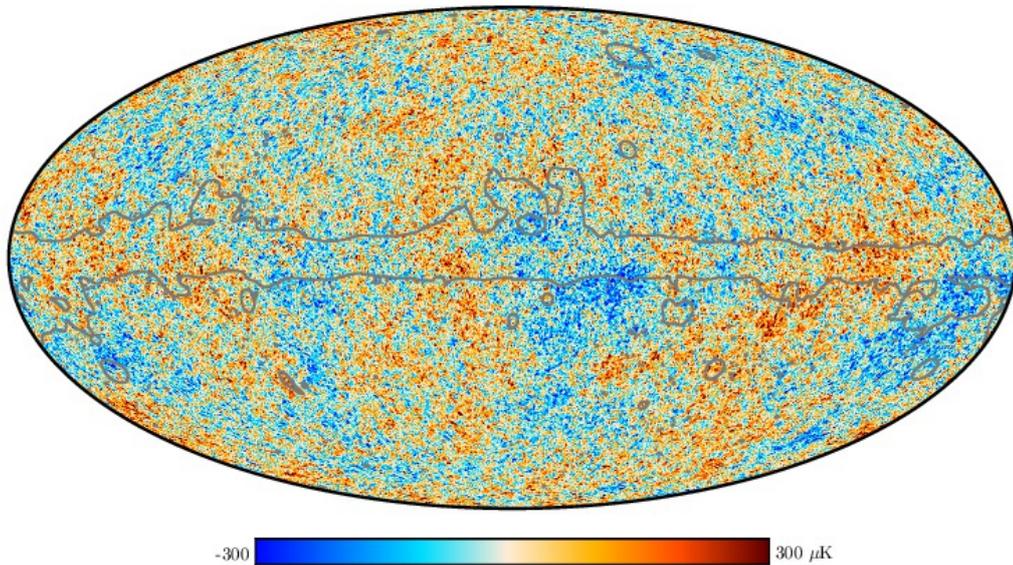


Figure 1.2. The colour-map of the temperature anisotropies obtained by the Planck collaboration in 2018. The picture shows the fluctuation around the mean black body temperature of the CMB.

The agreement between the power spectrum and the theoretical model is outstanding, as seen in Fig. 1.3. Each of the peaks and their overall amplitude and slope contain crucial information about the Universe.

In Fig. 1.3 the first peak act as a standard ruler since it is linked to the angular scale of the sound horizon at recombination. The first peak is influenced by the curvature of the Universe, and observation strongly suggest a spatially flat Universe. The relative amplitude of the peaks, depend on the ratio between Ω_m and Ω_r , the latter fixed by the temperature, so essentially the relative height of the peaks tracks the matter energy density. The last piece of information that relevant for this work is that the overall slope of the angular power spectrum depends on $\Omega_m h^2$ thus resulting in an estimation of H_0 .

This way of measuring H_0 depends on the chosen cosmological model, so is a model dependent measure, in this case the Λ CDM one.

Alongside the CMB, the Baryonic Acoustic Oscillation (BAO) phenomenon is pivotal. BAO arise from acoustic waves that propagated through the primordial plasma of baryons (protons and neutrons) and photons in the early universe [59]. During this epoch, baryons and photons were tightly coupled, and the pressure from photons and the gravitational pull from matter created oscillations, analogous to sound waves in a fluid. These oscillations were imprinted on the distribution of matter in the universe when it was still relatively young.

As the universe expanded, these acoustic waves left characteristic imprints on the density distribution of matter. These imprints are manifested as a periodic modulation in the spatial distribution of galaxies and other large-scale structures [60].

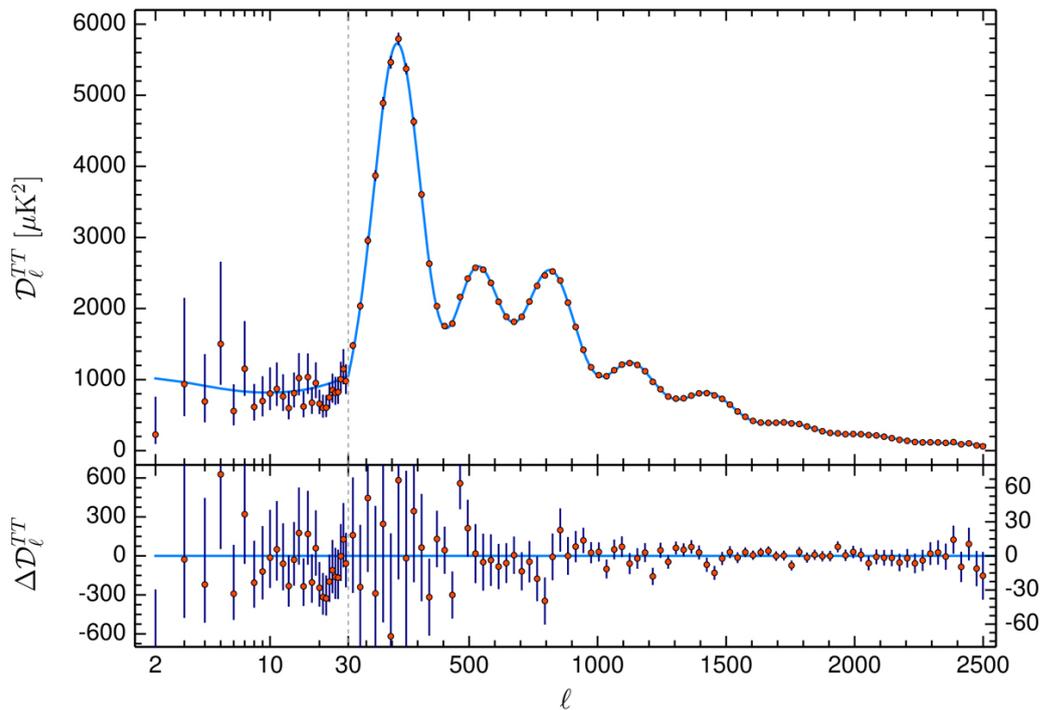


Figure 1.3. Planck's data points of CMB angular power spectrum of temperature anisotropies against the Λ CDM model, represented by the continuous curve. The bottom panel shows the residuals of the top panel.

The scale of these imprints, represents the maximum distance a sound wave could travel before recombination, the epoch when the universe became transparent to photons.

The characteristic scale of these acoustic oscillations, $\mathcal{O}(150)$ Mpc, serves as a standard ruler for measuring cosmic distances. By studying the statistical distribution of galaxies and measuring the correlation function or power spectrum, astronomers can detect the preferred separation scale associated with BAO.

BAO measurements provide a unique cosmological probe [61], offering constraints on fundamental parameters such as the geometry and expansion rate of the universe.

These spherical oscillations usually overlap each other, since the projection of those bubbles is observed on the celestial vault. Nonetheless, in a recent paper [62], a single BAO is visible in the overdensity of galaxy distribution.

1.2.2 Hubble Constant from Local Distance Indicators

The gold standard method of measuring H_0 in the late Universe, i.e., locally, is using either Cepheids or Type Ia Supernovae. Cepheids and Type Ia Supernovae, both mentioned before, are crucial to build the so-called "distance ladder". The distance ladder is a series of distance measurements from very close objects up to extra-galactic scale, in which each of the steps calibrate the next.

The first step is to measure the distance of a close object using, for example, the

parallax method. Then one can "climb" the ladder, using other strategies, one of which, is using Cepheids absolute magnitude from which the luminosity distance can be inferred via eq. (1.36)

$$M = m - 5 \log_{10} \frac{Dl}{10}. \quad (1.36)$$

Thanks to the satellite Gaia, the distance of Cepheids can be now obtained via parallax, improving the overall calibration of the ladder [63]. The general idea of the cosmic ladder is represented in Fig. 1.4. It is worth highlighting that this technique does not require a cosmological model. Once Cepheids are no longer visible, the focus

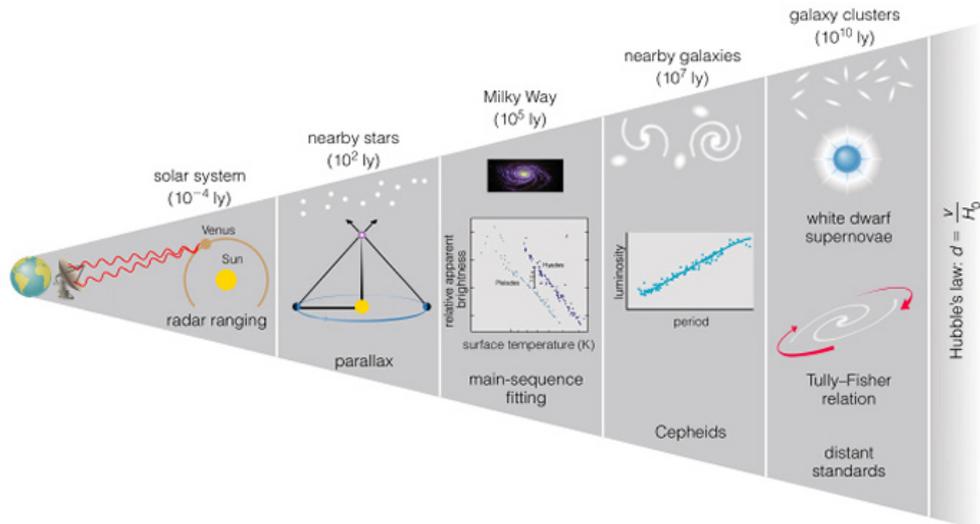


Figure 1.4. A representation of the various steps of the distance ladder. Image credit to David Darling.

shifts to Supernovae (SNe). As mentioned earlier, interest lies only in those SNe that can be "standardised", so attention turns to Type Ia SNe. After the critical mass is reached, a chain of nuclear reactions inside the white dwarf produces radioactive nuclei. The decay of these elements has a particular profile from which the intrinsic luminosity can be inferred.

The hierarchy of these measurements is evident in Fig. 1.5

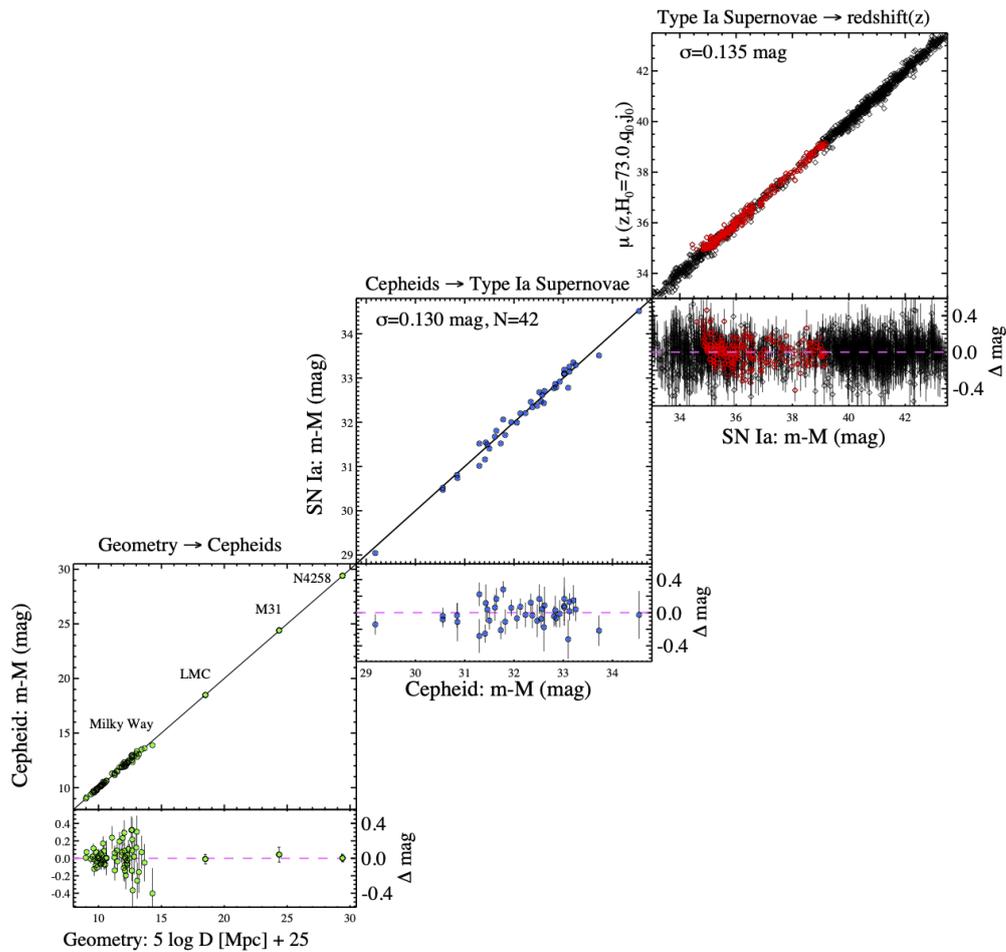


Figure 1.5. The complete distance ladder obtained by [64]. The image shows the hierarchical technique involved in the distance ladder method.

and once cosmological distances are reached, such that the redshift is cosmological, the Hubble constant is measured using the Hubble law, since for $z \ll 1$ eq. (1.31) reduces to $Dl = cz/H_0$.

The ever present challenge in this type of measurement is to reduce the systematic error introduced by this "step-by-step" technique. In this context, the work in [65] assesses the robustness of those type of data together with [66, 67]. In those works, the agreement with the early Universe estimation of H_0 do not increase much, even with new systematic or new Dark Energy models.

1.2.3 Two Headed Constant

The two main methods of measuring H_0 have been explained in the previous sections. A cosmological model that aims to explain the evolution of the Universe cannot admit the presence of two different H_0 values. In the era of Precision Cosmology, as visible in Fig. 1.6, advancements in analysis techniques for sources related to the local Universe lead to a sharp contrast between early and local Universe measurements.

Up to date, there are two well non-consistent values from [17, 16].

$$H_0^{early} = (67.66 \pm 0.62) \text{ km s}^{-1} \text{ Mpc}^{-1},$$

$$H_0^{local} = (73.29 \pm 0.90) \text{ km s}^{-1} \text{ Mpc}^{-1}.$$

It is important to emphasise that both measurements can be affected by systematic errors. From the distance ladder perspective, the most probable causes of error are either some unaccounted selection biases or systematics in the calibration of each step. On the other hand, in the Planck experiment, there is the A_{lens} parameter that represents the not well constrained amount of lensing in the tested model. Although these effects can mitigate the statistical tension, the different datasets tested and the sheer number of collaborations working on this problem make it unlikely that the two measurements will be compatible.

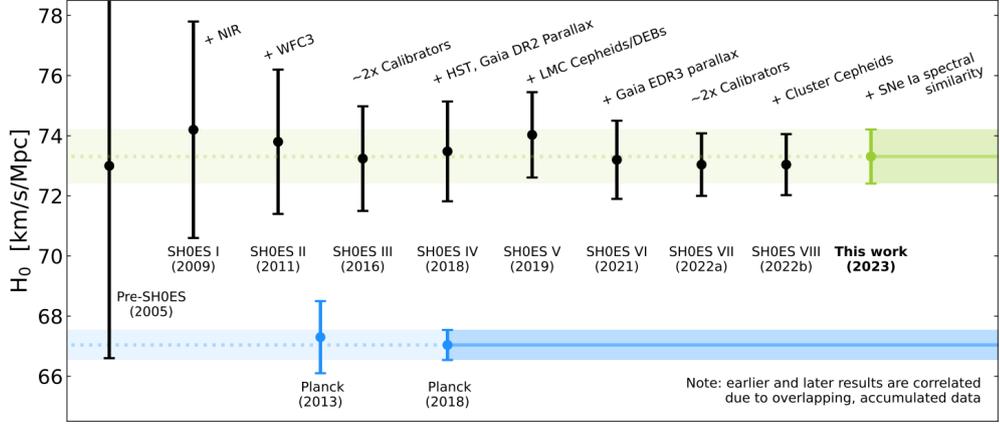


Figure 1.6. The evolution of the measurements of the Hubble constant using CMB and the distance ladder method, in light of new data analysis techniques. The Planck estimation is without BAO contribution. Ref [16]

Chapter 2

Gravitational Waves

GWs are a prediction of General Relativity, carrying information about the dynamical evolution of the gravitational field. The existence of GWs can be intuitively introduced by drawing an analogy with the electromagnetic case. Just as the perturbation of an electromagnetic field is carried by electromagnetic waves that spread information about the variation of the field, when there is a mass-energy distribution, it can be expected that information about its variation is transported by some kind of wave. Due to the nature of the metric tensor field $g_{\mu\nu}$, GWs must be understood as "metric waves" which, during their propagation, are responsible for the variation of the proper distance between space-time points. It is important to stress that, unlike electromagnetic waves, GWs are nonlinear, a result of the non-linearity, specifically the sesquilinearity, of the Einstein equations.

The process of how GWs arise from Einstein's field equations will be shown, followed by an exploration of some features of the waves, including deriving explicit formulae for some simple cases. The quadrupolar nature of GWs and the relevant parameters for describing GWs emission in a binary system will then be examined.

The field equations are written again, this time including the c^4 terms to provide an intuitive grasp of the magnitude of a gravitational wave. Einstein's field equations are

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \quad (2.1)$$

2.1 Deriving the Wave equation

To derive exact solutions, one must solve the field equations, describing both the source and the wave. Exact solutions require assuming some symmetry. Here, the perturbative approach is considered, and the metric tensor $g_{\mu\nu}$ is separated into a known background part, $\xi_{\mu\nu}$, and a perturbation, denoted by $h_{\mu\nu}$. It is required that, in some reference frame, the perturbation is small compared to $\xi_{\mu\nu}$, i.e.,

$$|h_{\mu\nu}| \ll |\xi_{\mu\nu}|. \quad (2.2)$$

The inverse metric is needed for calculations, and since the focus is on terms of the first order in the perturbation, the indices of h will be raised with the unperturbed

metric. With this in mind, the inverse metric is

$$g^{\mu\nu} = \xi^{\mu\nu} + h^{\mu\nu} + O(h^2). \quad (2.3)$$

To recast eq. (2.1) in a useful form, first substitute eq. (1.16) into eq. (2.1)

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu}, \quad (2.4)$$

then contract it with $g^{\mu\nu}$, obtaining

$$g^{\mu\nu}R_{\mu\nu} - \frac{1}{2}g^{\mu\nu}g_{\mu\nu}R = \frac{8\pi G}{c^4}g^{\mu\nu}T_{\mu\nu} \quad (2.5)$$

$$R - \frac{1}{2}R = \frac{8\pi G}{c^4}T^\alpha{}_\alpha \quad (2.6)$$

$$\frac{1}{2}R = \frac{8\pi G}{c^4}T^\alpha{}_\alpha. \quad (2.7)$$

This is used to rewrite eq. (2.1) as

$$R_{\mu\nu} = \frac{8\pi G}{c^4} \left(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T^\alpha{}_\alpha \right). \quad (2.8)$$

Then express the Ricci tensor in terms of the affine connection, compute the affine connection for the metric $g_{\mu\nu} = \xi_{\mu\nu} + h_{\mu\nu}$, and keep only the terms of order $\mathcal{O}(h)$. Remembering eq. (1.18), the affine connection in terms of the perturbed metric reads:

$$\begin{aligned} \Gamma^\alpha{}_{\mu\nu} &= \frac{1}{2}(\xi^{\alpha\gamma} - h^{\alpha\gamma}) \left[\left(\frac{\partial \xi_{\gamma\nu}}{\partial x^\mu} + \frac{\partial \xi_{\mu\gamma}}{\partial x^\nu} - \frac{\partial \xi_{\mu\nu}}{\partial x^\gamma} \right) + \left(\frac{\partial h_{\gamma\nu}}{\partial x^\mu} + \frac{\partial h_{\mu\gamma}}{\partial x^\nu} - \frac{\partial h_{\mu\nu}}{\partial x^\gamma} \right) \right] \\ &= \frac{1}{2}\xi^{\alpha\gamma} \left(\frac{\partial \xi_{\gamma\nu}}{\partial x^\mu} + \frac{\partial \xi_{\mu\gamma}}{\partial x^\nu} - \frac{\partial \xi_{\mu\nu}}{\partial x^\gamma} \right) + \frac{1}{2}\xi^{\alpha\gamma} \left(\frac{\partial h_{\gamma\nu}}{\partial x^\mu} + \frac{\partial h_{\mu\gamma}}{\partial x^\nu} - \frac{\partial h_{\mu\nu}}{\partial x^\gamma} \right) \\ &\quad - \frac{1}{2}h^{\alpha\gamma} \left(\frac{\partial \xi_{\gamma\nu}}{\partial x^\mu} + \frac{\partial \xi_{\mu\gamma}}{\partial x^\nu} - \frac{\partial \xi_{\mu\nu}}{\partial x^\gamma} \right) \\ &= \Gamma^\alpha{}_{\mu\nu}(\xi) + \Gamma^\alpha{}_{\mu\nu}(h) + O(h^2), \end{aligned} \quad (2.9)$$

where $\Gamma^\alpha{}_{\mu\nu}(h)$ collects all terms linear in h in eq. (2.9).

equation (2.9) can be used to calculate the Ricci tensor, with the understanding that only terms up to first order in h are needed. The right-hand side of equation (2.8) is expressed as,

$$\begin{aligned} R_{\mu\nu} &= R_{\mu\nu}(\xi) + \frac{\partial \tilde{\Gamma}^\alpha{}_{\mu\nu}}{\partial x^\alpha} - \frac{\partial \tilde{\Gamma}^\alpha{}_{\mu\alpha}}{\partial x^\nu} + \Gamma^\alpha{}_{\sigma\alpha} \tilde{\Gamma}^\sigma{}_{\mu\nu} + \Gamma^\alpha{}_{\sigma\nu} \tilde{\Gamma}^\sigma{}_{\mu\alpha} \\ &\quad + \tilde{\Gamma}^\alpha{}_{\sigma\alpha} \Gamma^\sigma{}_{\mu\nu} + \tilde{\Gamma}^\alpha{}_{\sigma\nu} \Gamma^\sigma{}_{\mu\alpha}, \end{aligned} \quad (2.10)$$

where the tilde denotes terms in h .

For simplicity, the background metric $\xi_{\mu\nu}$ is substituted with the flat metric $\eta_{\mu\nu}$. Consequently, the affine connections without a tilde in equation (2.10) vanish, as all the Christoffel symbols for a flat metric are zero. With $\xi_{\mu\nu}$ now flat, equation (1.18)

becomes

$$\Gamma^{\alpha}_{\mu\nu} = \frac{1}{2}\eta^{\alpha\gamma} \left(\frac{\partial h_{\gamma\nu}}{\partial x^{\mu}} + \frac{\partial h_{\mu\gamma}}{\partial x^{\nu}} - \frac{\partial h_{\mu\nu}}{\partial x^{\gamma}} \right). \quad (2.11)$$

In equation (2.10), only the derivatives are dealt with. Additionally, since equation (2.10) expresses the left-hand side of the Einstein field equations, the term $R_{\mu\nu}(\xi)$ simplifies with the unperturbed part of the right side. This allows the writing of

$$\square_F h_{\mu\nu} - \left[\frac{\partial^2 h^{\sigma}_{\nu}}{\partial x^{\sigma} \partial x^{\mu}} + \frac{\partial^2 h^{\sigma}_{\mu}}{\partial x^{\sigma} \partial x^{\nu}} - \frac{\partial^2 h^{\sigma}_{\sigma}}{\partial x^{\mu} \partial x^{\nu}} \right] = -\frac{16\pi G}{c^4} \left(\tilde{T}_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} \tilde{T}^{\sigma}_{\sigma} \right), \quad (2.12)$$

where \tilde{T} represents the perturbation of the stress-energy tensor, and \square_F is the D'Alembertian in flat spacetime $\square_F = \eta^{\mu\nu} \frac{\partial}{\partial x^{\mu}} \frac{\partial}{\partial x^{\nu}}$.

It is emphasized that these equations are not uniquely determined, and a coordinate transformation can always be performed in such a way that eq. (2.12) still holds. A particular transformation that simplifies the above equation is the harmonic gauge condition, namely

$$g^{\mu\nu} \Gamma^{\sigma}_{\mu\nu} = 0. \quad (2.13)$$

Rewriting the condition in equation (2.13) in terms of the metric perturbation yields:

$$\begin{aligned} 0 &= g^{\mu\nu} \Gamma^{\sigma}_{\mu\nu} = \eta^{\mu\nu} \Gamma^{\sigma}_{\mu\nu} = \frac{1}{2} \eta^{\mu\nu} \eta^{\sigma\alpha} (\partial_{\mu} h_{\alpha\nu} + \partial_{\nu} h_{\mu\alpha} - \partial_{\alpha} h_{\mu\nu}) \\ &= \frac{1}{2} \eta^{\sigma\alpha} (\partial_{\mu} h^{\mu}_{\alpha} + \partial_{\nu} h^{\nu}_{\alpha} - \partial_{\alpha} h^{\nu}_{\nu}) \\ &= \frac{1}{2} \eta^{\sigma\alpha} (2\partial_{\mu} h^{\mu}_{\alpha} - \partial_{\alpha} h^{\nu}_{\nu}). \end{aligned} \quad (2.14)$$

Analysing the terms inside the brackets, the harmonic gauge condition translates into

$$\frac{\partial h^{\mu}_{\nu}}{\partial x^{\mu}} - \frac{1}{2} \frac{\partial h^{\mu}_{\mu}}{\partial x^{\nu}} = 0, \quad (2.15)$$

enabling the system to be written as:

$$\begin{cases} \square_F h_{\mu\nu} = -\frac{16\pi G}{c^4} \left(\tilde{T}_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} \tilde{T} \right), \\ \frac{\partial h^{\mu}_{\nu}}{\partial x^{\mu}} - \frac{1}{2} \frac{\partial h^{\mu}_{\mu}}{\partial x^{\nu}} = 0. \end{cases} \quad (2.16)$$

The system can then be recast using the reverse procedure that transformed eq. (2.1) into eq. (2.8). For simplicity, the tilde on the stress-energy tensor is omitted, as the background part of the Einstein field equation is satisfied when writing eq. (2.16). Thus, in this context, the stress-energy tensor only sources the perturbed part of the metric.

To simplify eq. (2.16), it is noted that the perturbed part of the Ricci tensor reads $\square_F h_{\mu\nu}$, which is the only term in eq. (2.12) surviving the gauge condition. Using eq. (2.7), it is found that

$$\frac{8\pi G}{c^4} \eta_{\mu\nu} T = \frac{1}{2} \eta_{\mu\nu} R = \frac{1}{2} \eta_{\mu\nu} \square_F h^{\sigma}_{\sigma}. \quad (2.17)$$

Introducing a new tensor $\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h$, eq. (2.16) is rewritten as

$$\begin{cases} \square_F \bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4} T_{\mu\nu}, \\ \frac{\partial}{\partial x^\mu} \bar{h}^\mu{}_\nu = 0. \end{cases} \quad (2.18)$$

In vacuum, eq. (2.16) becomes

$$\begin{cases} \square_F \bar{h}_{\mu\nu} = 0, \\ \frac{\partial}{\partial x^\mu} \bar{h}^\mu{}_\nu = 0. \end{cases} \quad (2.19)$$

It is evident from these expressions that metric perturbations satisfy the wave equations, thus propagating as waves at the speed of light.

Concluding this section, two important properties of GWs are demonstrated using plane wave solutions. A plane wave, being the simplest solution of the wave equation, is given by

$$\bar{h}_{\mu\nu} = \Re[A_{\mu\nu} e^{ik_\sigma x^\sigma}], \quad (2.20)$$

where A is the polarisation tensor and k is the wave vector such that $k_\alpha = (\omega/c, \mathbf{k})$.

Using a plane wave in eq. (2.19), it is seen that the second equation leads to the conclusion that GWs are transverse waves, with the polarisation tensor being orthogonal to the wave vector:

$$\frac{\partial \bar{h}^\mu{}_\nu}{\partial x^\mu} = \frac{\partial \eta^{\mu\sigma} \bar{h}_{\sigma\nu}}{\partial x^\mu} = \frac{\partial \eta^{\mu\sigma} A_{\sigma\nu} e^{ik_\alpha x^\alpha}}{\partial x^\mu} = \eta^{\mu\sigma} A_{\sigma\nu} k_\mu = k_\mu A^\mu{}_\nu = 0. \quad (2.21)$$

A gravitational plane wave thus has a null wave vector, implying that both the group and phase velocities of GWs are equal to the speed of light.

2.2 Gauge Invariance and the Traceless Transverse Gauge

General relativity is invariant under arbitrary coordinate transformations of the form

$$x'^\mu = x^\mu + \chi^\mu(x), \quad (2.22)$$

where $'$ refers to the "new" coordinate system. The theory is thus invariant under diffeomorphism, and this invariance can be utilized to further reduce the degrees of freedom of the $\bar{h}_{\mu\nu}$, commonly referred to as the gauge invariance of General Relativity.

In principle, $g_{\mu\nu}$ has 16 degrees of freedom. However, as it is a metric tensor, it must be symmetric, leaving only 10 independent components. The harmonic gauge utilises four degrees of freedom, leaving six to work with. The invariance under diffeomorphism is then used, ultimately leaving only two independent components for the perturbation $h_{\mu\nu}$.

A general metric tensor transforms under diffeomorphism as follows:

$$g'_{\mu\nu} = \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial x'^\nu} g_{\alpha\beta}. \quad (2.23)$$

Using eq. (2.22) in eq. (2.23) for a perturbation over a flat space, the new metric tensor reads

$$g'_{\mu\nu} = (\delta_{\mu}^{\alpha} + \partial_{\mu}\chi^{\alpha})(\delta_{\nu}^{\beta} + \partial_{\nu}\chi^{\beta})(\eta_{\alpha\beta} + h_{\alpha\beta}), \quad (2.24)$$

using ∂_{μ} to indicate a partial derivative with respect to x^{μ} . Keeping in mind that only terms up to first order in h are of interest, and χ is of the same order as $h_{\mu\nu}$, eq. (2.24) reduces to

$$g'_{\mu\nu} \approx \eta_{\mu\nu} + h_{\mu\nu} - \partial_{\mu}\chi_{\nu} - \partial_{\nu}\chi_{\mu} = \eta_{\mu\nu} + h'_{\mu\nu}, \quad (2.25)$$

with the implicit definition $h'_{\mu\nu} = h_{\mu\nu} - \partial_{\mu}\chi_{\nu} - \partial_{\nu}\chi_{\mu}$. This forms the basis for further simplification of the perturbation.

Since only the coordinates are being acted upon, simply changing how they are described, if h is a solution of a physical problem, so is h' and vice versa. Recalling the definition of \bar{h} ,

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h, \quad (2.26)$$

some condition can be imposed on \bar{h}' ,

$$\begin{aligned} \bar{h}'_{\mu\nu} &= h'_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h' \\ &= h_{\mu\nu} - (\partial_{\mu}\chi_{\nu} + \partial_{\nu}\chi_{\mu}) - \frac{1}{2}\eta_{\mu\nu}(h - 2\partial_{\rho}\chi^{\rho}) \\ &= \bar{h}_{\mu\nu} - (\partial_{\mu}\chi_{\nu} + \partial_{\nu}\chi_{\mu} - \eta_{\mu\nu}\partial_{\rho}\chi^{\rho}) \\ &= \bar{h}_{\mu\nu} - \Psi_{\mu\nu}. \end{aligned} \quad (2.27)$$

It can be shown that if χ_{μ} is chosen such that $\square_F\chi_{\mu} = 0$, then $\bar{h}'_{\mu\nu}$ can be made traceless and $\bar{h}'_{0i} = 0$. This is called the Transverse Traceless representation h^{TT} , the explicit form, for a GW propagating along the x^3 direction, is ($x^{\alpha} = (ct, 0, 0, z)$)

$$h^{\text{TT}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_{+} & h_{\times} & 0 \\ 0 & h_{\times} & -h_{+} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \cos[\omega(t - z/c)], \quad (2.28)$$

where h_{+} and h_{\times} are the components of the polarisation vector $A_{\mu\nu}$ of the GW.

For a wave travelling in the general direction $\hat{\mathbf{n}}$, the projector P_{ij} can be introduced, defined as

$$P_{ij}(\hat{\mathbf{n}}) = \delta_{ij} - n_i n_j. \quad (2.29)$$

With this projector, a Λ matrix can be constructed that can transform a $h_{\mu\nu}$ computed in the harmonic gauge to one that is transverse and traceless. The matrix is defined as

$$\Lambda_{ij,kl} = P_{ik}P_{jl} - \frac{1}{2}P_{ik}P_{jl}, \quad (2.30)$$

and the **TT** gauge can be obtained by

$$h_{ij}^{\mathbf{TT}} = \Lambda_{ij,kl} h_{kl}, \quad (2.31)$$

where h_{kl} was in the harmonic gauge.

2.3 Solutions In The Low Velocity Regime

In the presence of a non-zero stress-energy tensor, solutions of eq. (2.12) are sought under the *slow motion approximation*. This approximation assumes that the physical dimension of the source is much smaller than the wavelength of the GW, meaning the source extends only in a small region of typical dimension ϵ .

$$|x^i| < \epsilon \rightarrow T_{\mu\nu} \neq 0, \quad (2.32)$$

so that the wavelength of the GW, $\lambda = 2\pi c/\omega$, satisfies

$$\frac{2\pi c}{\omega} \gg \epsilon \rightarrow \epsilon\omega \ll c \rightarrow c \gg v_{\text{typical}}. \quad (2.33)$$

This chain of implications states that the typical velocities of the physical processes of the system are small compared to the speed of light. It is worth noting that during a Black Hole merger, the two objects can move up to one third of the speed of light, thus breaking this assumption.

Moving to the first equation in eq. (2.12), it can be solved using Green's method, in other words, by finding a function that satisfies

$$\square_x G(x - x') = \delta^4(x - x'), \quad (2.34)$$

with δ^4 being the four-dimensional Dirac delta and $\square = \square_F$. Using the Green's function, explicit solutions for $\bar{h}_{\mu\nu}$ are

$$\bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4} \int_{\mathcal{V}} G(x - x') T_{\mu\nu}(x') d^4 x', \quad (2.35)$$

where \mathcal{V} is a four-dimensional region $V \times I$, where V is the boundary of the source outside which the stress-energy tensor is zero and I is a time interval.

Similar equations arise in radiative problems in electromagnetism. Although gravity is non-linear, since linear equations for \bar{h} are being solved, solutions from electromagnetism can be borrowed. The Green function can be shown to be

$$G(x - x') = -\frac{1}{4\pi|\mathbf{x} - \mathbf{x}'|} \delta(x_{\text{ret}}^0 - x'^0), \quad (2.36)$$

introducing the concept of *retarded* time $x_{\text{ret}}^0 = ct_{\text{ret}}$ and t_{ret} is

$$t_{\text{ret}} = t - \frac{|\mathbf{x} - \mathbf{x}'|}{c}. \quad (2.37)$$

The idea of the retarded time is that physical information cannot travel faster than light, so any solution can depend only on events in causally connected regions. Substituting the explicit expression of the Green function into the solution yields

$$\bar{h}_{\mu\nu} = \frac{4G}{c^4} \int_V \frac{1}{4\pi|\mathbf{x} - \mathbf{x}'|} T_{\mu\nu}(t_{\text{ret}}, \mathbf{x}') d^3x'. \quad (2.38)$$

To solve eq. (2.38), moving to Fourier space is convenient. Before doing that, using $|\mathbf{x}| = r \rightarrow |\mathbf{x} - \mathbf{x}'| = r - \mathbf{x}' \cdot \hat{\mathbf{n}} + \mathcal{O}(\epsilon^2/r)$ simplifies $|\mathbf{x} - \mathbf{x}'| \sim r$ and can be taken out of the integral. The explicit definition in terms of the ω variable is

$$T_{ij} \left(t - \frac{r}{c} + \frac{\mathbf{x}' \cdot \hat{\mathbf{n}}}{c}, \mathbf{x}' \right) = \int \frac{1}{(2\pi)^4} \tilde{T}_{ij}(\omega, \mathbf{k}) e^{-i\omega(t - \frac{r}{c} + \frac{\mathbf{x}' \cdot \hat{\mathbf{n}}}{c}) + i\mathbf{k} \cdot \mathbf{x}'}. \quad (2.39)$$

Using the slow-motion approximation, in the exponent the term $\omega\mathbf{x}'/c$ satisfies $\omega\mathbf{x}'/c \leq \omega\epsilon/c \sim v/c \ll 1$. This leads to

$$e^{-i\omega(t - \frac{r}{c} + \frac{\mathbf{x}' \cdot \hat{\mathbf{n}}}{c})} = e^{-i\omega(t - \frac{r}{c})} \left[1 - i\frac{\omega}{c} x^l n^l + \frac{1}{2} \left(-i\frac{\omega}{c}\right)^2 x^l x^m n^l n^m + \dots \right], \quad (2.40)$$

therefore

$$T_{ij} \left(t - \frac{r}{c} + \frac{\mathbf{x}' \cdot \hat{\mathbf{n}}}{c}, \mathbf{x}' \right) \approx T_{ij} \left(t - \frac{r}{c}, \mathbf{x}' \right) + \frac{x^l n^l}{c} \partial_0 T_{ij} + \frac{1}{2c^2} x^l x^m n^l n^m \partial_0^2 T_{ij} + \dots, \quad (2.41)$$

2.4 Momenta And Quadrupole

The concept of momenta of the stress-energy tensor, $S^{ij}(t)$, can now finally be introduced, which is

$$S^{ij,lmno\dots} = \int T^{ij}(t, \mathbf{x}) x^l x^m x^n x^o d^3x, \quad (2.42)$$

and at first order,

$$S^{ij} = \int T^{ij}(t, \mathbf{x}) d^3x. \quad (2.43)$$

By definition, the momenta are symmetrical in i and j since T^{ij} is symmetrical and invariant under permutation of the other indices, but $S^{ij,k} \neq S^{ik,j}$. Using eq. (2.42) to express each term in eq. (2.41), the main equation for GW emission, the multipolar expansion, can be written down:

$$h_{ij}^{TT}(t, \mathbf{x}) = \frac{4G}{rc^4} \Lambda_{ij,kl} \left[S^{kl} + \frac{1}{c} n_m \dot{S}^{kl,m} + \frac{1}{2c^2} n_m n_p \ddot{S}^{kl,mp} \right], \quad (2.44)$$

evaluated at the retarded time.

This equation is insightful, clearly showing the decay of the signal proportional to r^{-1} , where r is the distance from the source. The amplitude of the GW is then defined by matching the internal solution, the one just found, to the exterior solution,

i.e., when the stress-energy tensor vanishes. The nature of gravity implies that, unlike electromagnetism which has dipole radiation, the first term contributing to the signal in gravity is the quadrupole term in eq. (2.44).

The leading term in eq. (2.44) can be evaluated once it is recalled that $T^{\mu\nu}$ satisfies the conservation equation $\partial_\mu T^{\mu\nu} = 0$. Similarly to S^{ij} , the momenta for the 00 component of $T^{\mu\nu}$, M , and for the 0*i* components, P^i , can be defined. This allows for the expression of S^{ij} , and its higher orders, in terms of P^i , then the component $\nu = 0$ of the conservation equation can link M and P .¹

The goal is to express $S^{ij,kl\dots v}$ as a combination of $M^{ij\dots v}$ and $P^{i,j\dots v}$, and then show that the first contribution is the quadrupole one. The new momenta M and P are now defined as

$$\begin{aligned} M^{ij\dots v} &= \int T^{00}(t, \mathbf{x}) x^i x^j \dots x^v d^3x, \\ P^{i,jl\dots v} &= \int T^{0i}(t, \mathbf{x}) x^i x^j \dots x^v d^3x. \end{aligned} \quad (2.45)$$

It is straightforward to prove that $\dot{M} = c\partial_0 M = 0$ since

$$c\partial_0 M = c \int_V \partial_0 T^{00} d^3x = \underbrace{-c \int_V \partial_i T^{0i} d^3x}_{\text{conservation equation}} = \overbrace{-c \int_{\partial V} T^{0i} d\sigma^i}^{T^{\mu\nu}=0 \text{ on the boundary}} = 0, \quad (2.46)$$

and the same is true for \dot{P}^i . The conservation equation provides a link between M and P since

$$\dot{M}^i = c \int_V x^i \partial_0 T^{0i} d^3x = - \int_V x^i \partial_k T^{0k} d^3x = \int_V \partial_k x^i T^{0k} d^3x = P^i, \quad (2.47)$$

and similarly,

$$\dot{M}^{ij} = P^{i,j} + P^{j,i}. \quad (2.48)$$

The last equation to prove in order to obtain an expression for the leading order of eq. (2.44) is

$$\dot{P}^{i,j} = c \int \partial_0 T^{0i} x^j d^3x = - \int \partial_k T^{ki} x^j d^3x = \int T^{ki} \partial_k x^j d^3x = \int T^{ki} \delta_k^j d^3x = S^{ij}. \quad (2.49)$$

Using eq. (2.49) and eq. (2.48) in eq. (2.42), the result is

$$h_{ij}^{TT}(t, \mathbf{x}) = \frac{2G}{rc^4} \Lambda_{ij,kl} \left[\ddot{M}^{kl} \left(t - \frac{r}{c} \right) \right]. \quad (2.50)$$

Taking a moment to remark on some aspects of GWs, it is observed that the amplitude of the signal is proportional to $G/c^4 \sim 8 \times 10^{-50} \text{s}^2/\text{g}/\text{cm}$, implying that GWs are "weak", being a small perturbation of the background. Secondly, having

¹This is possible since $T^{\mu\nu}$ vanishes at the boundary of the source.

$\dot{P}^i = 0$, means that for a closed system of particles, the gravitational dipole moment

$$\mathbf{D}_g = \sum_c m_c \mathbf{r}_c \quad (2.51)$$

is conserved, so there is no dipole emission, as clearly stated in eq. (2.50). Moreover, a spherical axisymmetric object has a constant quadrupole moment, so a rotating axisymmetric body does not emit GWs. In order to radiate, a single body has to be asymmetric and spinning.²

The symmetric tensor M^{ij} can always be represented as the sum of a symmetric traceless tensor and its trace:

$$M^{ij} = \left(M^{ij} - \frac{1}{3} \delta^{ij} M_{kk} \right) + \frac{1}{3} \delta^{ij} M_{kk}. \quad (2.52)$$

The second term, the trace, can be disregarded when working in the TT gauge. Defining the traceless part Q^{ij} as the quadrupole, the expression for $h_{ij}^{TT}(t, \mathbf{x})$ becomes

$$h_{ij}^{TT}(t, \mathbf{x}) = \frac{2G}{rc^4} \Lambda_{ij,kl} \left[\ddot{Q}^{kl} \left(t - \frac{r}{c} \right) \right]. \quad (2.53)$$

Considering a GW propagating in the $\hat{\mathbf{z}}$ direction, the wave amplitude in the TT gauge can be computed with

$$\Lambda_{ij,kl} A_{kl} = \left(P_{ik} P_{jl} - \frac{1}{2} P_{ik} P_{jl} \right) A_{kl}. \quad (2.54)$$

The projector P must be of the form $diag(1, 1, 0)$, as GWs are transverse, leading to

$$\Lambda_{ij,kl} A_{kl} = \begin{pmatrix} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} \frac{A_{11}+A_{22}}{2} & A_{12} & 0 \\ A_{21} & \frac{A_{11}+A_{22}}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (2.55)$$

Identifying with the previously introduced two polarisation h_+ and h_x , the final derivations are

$$\begin{aligned} h_+ &= \frac{G}{rc^4} (\ddot{Q}_{11} - \ddot{Q}_{22}), \\ h_x &= \frac{2G}{rc^4} \ddot{Q}_{12}. \end{aligned} \quad (2.56)$$

2.5 Binary Systems

The study of a system of two objects orbiting their common centre of mass is now undertaken, ignoring any energy loss due to the emission of GWs for the moment. The energy carried away from the system, which shrinks the orbit in a positive feedback loop, will eventually lead to the merge of the two objects. This phenomenon, first observed by Hulse and Taylor [68] in 1975, was an initial indirect proof of the

²This is how we know that Neutron Stars are quite smooth, since there are no detection of GWs from a single fast-spinning Neutron Star.

existence of GWs. Simplifying the problem, the two masses are considered point-like, with masses m_1 and m_2 , respectively. The orbital separation is denoted as l_0 , and the vectors r_1 and r_2 point at the masses in the reference frame, as shown in Fig. 2.1. The total mass M and the reduced mass are defined as,

$$\mu = \frac{m_1 m_2}{M}. \quad (2.57)$$

Assuming planar motion, the system can be set in the (x, y) -plane without loss of generality. In this frame of reference, the relationships are

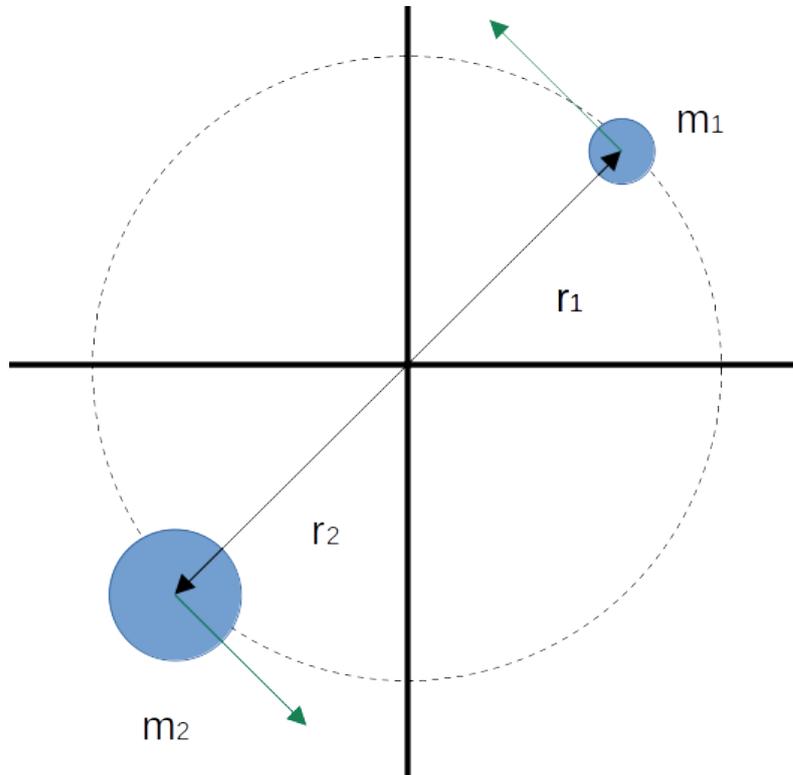


Figure 2.1. A schematic representation of the system.

$$l_0 = r_1 + r_2, \quad r_1 = \frac{m_2 l_0}{M}, \quad r_2 = \frac{m_1 l_0}{M}, \quad (2.58)$$

and the orbital frequency ω_o , according to Kepler's law of motion, is

$$\omega_o = \sqrt{\frac{GM}{l_0^3}}. \quad (2.59)$$

The quadrupole moment is then computed. On the orbital plane, the coordinates for the masses are:

$$\begin{aligned} x_1 &= \frac{m_2}{M} l_0 \cos \omega_o t, & y_1 &= \frac{m_2}{M} l_0 \sin \omega_o t, \\ x_2 &= \frac{m_1}{M} l_0 \cos \omega_o t, & y_2 &= \frac{m_1}{M} l_0 \sin \omega_o t, \end{aligned} \quad (2.60)$$

and the stress-energy tensor for an ensemble of point-like masses is

$$T^{\mu\nu} = \sum_i \gamma_i m_i \frac{dx_i^\mu}{dt} \frac{dx_i^\nu}{dt} \delta^3(\mathbf{x} - \mathbf{x}_i). \quad (2.61)$$

For a non-relativistic system, the 00-component of $T^{\mu\nu}$ is

$$T^{00} = c^2 \sum_{i=1}^2 m_i \delta(x - x_i) \delta(y - y_i). \quad (2.62)$$

With these components, the quadrupole can be computed:

$$\begin{aligned} M_{xx} &= m_1 \int_V x^2 \delta(x - x_1) \delta(y - y_1) \delta(z) dx dy dz \\ &+ m_2 \int_V x^2 \delta(x - x_2) \delta(y - y_2) \delta(z) dx dy dz \\ &= \frac{\mu}{2} l_0^2 \cos 2\omega_o t + C_0, \end{aligned} \quad (2.63)$$

where C_0 is a constant. Similarly, for q_{yy} and q_{xy} , the results are

$$\begin{aligned} M_{xx} &= \frac{\mu}{2} l_0^2 \cos 2\omega_o t + C_0, \\ M_{yy} &= -\frac{\mu}{2} l_0^2 \cos 2\omega_o t + C_1, \\ M_{xy} &= \frac{\mu}{2} l_0^2 \sin 2\omega_o t. \end{aligned} \quad (2.64)$$

Finally, the signal in the TT gauge is described as follows:

$$A_{ij} = \begin{pmatrix} \cos 2\omega_o t & \sin 2\omega_o t & 0 \\ \sin 2\omega_o t & -\cos 2\omega_o t & 0 \\ 0 & 0 & 0 \end{pmatrix} A_0, \quad (2.65)$$

where A_0 is defined as

$$A_0 = \frac{4\mu M G^2}{l_0 c^4}. \quad (2.66)$$

Thus, the components of the gravitational wave in the TT gauge are given by

$$\begin{aligned} h_{xx}^{TT} &= -\frac{A_0}{z} \cos 2\omega_o \left(t - \frac{z}{c}\right), \\ h_{yy}^{TT} &= -h_{xx}^{TT}, \\ h_{xy}^{TT} &= -\frac{A_0}{z} \sin 2\omega_o \left(t - \frac{z}{c}\right). \end{aligned} \quad (2.67)$$

The term ω_o^2 in the final equation, arising from the time derivatives, can be related to the orbital frequency via eq. (2.59). It is noted that the main spectral contribution comes at twice the orbital frequency, and the frequency increases as the orbit shrinks due to radiation.

2.6 Energy of GWs

GWs carry energy away from the system, sourced from the orbital energy E_o , comprising potential and kinetic energy:

$$E_o = E_k + E_{pot} = -\frac{Gm_1m_2}{2l_0}. \quad (2.68)$$

The power radiated away is described by

$$P = \frac{32c^5}{5G} \left(\frac{GM_c \omega_{gw}}{2c^3} \right)^{10/3}, \quad (2.69)$$

with the chirp mass $M_c = \mu^{3/5} M^{2/5}$ defined. In the 'quasi-circular' motion, where $\dot{\omega}_o \ll \omega_o^2$, the decrease in radial separation \dot{l}_0 is small:

$$\dot{l}_0 = -\frac{2l_0}{3} \frac{\dot{\omega}_o}{\omega_o} = -\frac{2l_0\omega_o}{3} \frac{\dot{\omega}_o}{\omega_o^2}. \quad (2.70)$$

The time derivative of the orbital frequency, $\dot{\omega}_{gw}$, is derived by equating the emitted power to the variation in orbital energy, linking ω_o to the orbital energy using eq. (2.59):

$$\omega_{gw} = \frac{12\sqrt[3]{2}}{5} \left(\frac{GM_c}{c^3} \right)^{5/3} \omega_{gw}^{11/3}, \quad (2.71)$$

or in terms of frequencies:

$$f_{gw} = \frac{96\sqrt[3]{\pi^8}}{5} \left(\frac{GM_c}{c^3} \right)^{5/3} f_{gw}^{11/3}. \quad (2.72)$$

Defining the time of coalescence t_{coal} as the time of the bodies' merge, and using the variable τ , the time-dependent frequency of the GW signal is integrated from eq. (2.72) to obtain

$$f(\tau) = \frac{1}{\pi} \left(\frac{GM_c}{c^3} \right)^{-5/8} \left(\frac{5}{256\tau} \right)^{3/8}. \quad (2.73)$$

In conclusion, it is emphasised that the above equations are valid under specific limits and constraints discussed above.

The eq. (2.73) diverges in a finite time when $\tau = 0$. This indicates that the model breaks down before the coalescence time. Firstly, the slow motion approximation will eventually fail, as ω_o increases with time. During coalescence, the orbit shrinks, and the point-like approximation for the bodies becomes invalid. All calculations above assume a flat background, which does not hold for BH systems due to unstable orbits.³ Tidal forces, especially in Neutron Star (NS) and BH or NS-NS mergers, can accelerate the decay of the orbit.

2.7 Propagation

With expressions for the amplitudes of the two polarisation of GWs in flat space derived, attention turns to sources at cosmological distances. Advanced gravitational-wave detectors such as ET and LISA may detect signals from binaries at these distances [33]. Gravitational Waves travel along null geodesics of spacetime. Gravity, being sourced by both mass and energy, implies that GWs, carrying energy, have a form of auto-interaction, though it is negligible. The weak field approximation ensures the GW does not alter the background metric. Hence, the FRLW metric remains unchanged, and interaction between polarisation is negligible.

This section aims to understand how signals from cosmologically distant binary merges are influenced by the Universe's expansion. The concept of the local wave zone is introduced, defined as a region where the distance to the source is large enough to exhibit $1/r$ wave-like behaviour while manumitting the impact of the Universe's expansion. Physical distances within the wave zone are expressed as

$$r_{phys} = a(t_e)dc, \quad (2.74)$$

with t_e being the time of emission and dc the comoving distance.

The main consideration is to include cosmological redshift as with photons. The affected quantities are:

$$\begin{aligned} \lambda^{obs} &= (1+z)\lambda^{emiss}, \\ dt^{obs} &= (1+z)dt^{emiss}, \\ f^{obs} &= \frac{1}{1+z}f^{emiss}. \end{aligned} \quad (2.75)$$

The GW amplitude equations are expressed in terms of frequency using eq. (2.59) and $\omega = 2\pi f$. Retarded time t_s^{ret} in the source frame and other quantities are appropriately marked with subscript s . The GW frequency relation is:

$$h(t_s^{ret}) = \frac{4}{a(t_e)dc} \left(\frac{GM_c}{c^2} \right)^{5/3} \left(\frac{\pi f_s(t_s^{ret})}{c} \right)^{2/3}, \quad (2.76)$$

³Once the Innermost Stable Circular Orbit (ISCO) is passed, the idea of a slowly shrinking circular orbit is not feasible.

with the frequency satisfying

$$f_s = \frac{1}{\pi} \left(\frac{GM_c}{c^3} \right)^{-5/8} \left(\frac{5}{256\tau_s} \right)^{3/8}. \quad (2.77)$$

Using eq. (2.75), the equation is recast in the observer's frame, with $Dl = (1+z)dc$, leading to

$$h(t_{obs}^{ret}) = \frac{4}{Dl} \left(\frac{G(1+z)M_c}{c^2} \right)^{5/3} \left(\frac{\pi f_s(t_{obs}^{ret})}{c} \right)^{2/3}, \quad (2.78)$$

and introducing the redshifted chirp mass $\mathcal{M}_c = (1+z)M_c$, the final formula is

$$h(t_{obs}^{ret}) = \frac{4}{Dl} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{5/3} \left(\frac{\pi f_s(t_{obs}^{ret})}{c} \right)^{2/3}. \quad (2.79)$$

The functional form remains the same as before, using the luminosity distance and the redshifted chirp mass.

Chapter 3

Poking Around The Dark

In this chapter, the working principle of a ground-based interferometer is introduced, focusing on the Einstein Telescope (ET) case, and then a brief introduction to the Euclid mission is given, with focus on the spectroscopic sample.

In the previous section, it was established that $h(t)$ generally comprises two polarisation with amplitudes given by eq. (2.79). By analysing the signal from each polarisation, the orbital inclination relative to the line of sight can be determined. With measurements of the frequency and its time derivative, the redshifted chirp mass can be evaluated and the luminosity distance can be calculated. This provides a general overview of how D_l can be measured; the practical implementation will be discussed in the dedicated section.

3.1 Gravitational Waves Detectors

This section describes the general properties of a ground-based interferometer for detecting GW signals, then focuses on the specific case of ET.

The earliest efforts to detect GWs utilised Resonant Mass detectors. These detectors, typically incorporating massive resonant bars, are based on the principle that a passing GW induces small length oscillations in a massive object, generating an observable signal.

Advancements in detector technologies led to the development of interferometric detectors, with the first direct detection made in 2015 [69]. These detectors, such as LIGO and Virgo, use laser interferometry to measure tiny changes in length caused by passing GWs.

The upcoming ET interferometer aims to significantly enhance our ability to observe and understand the universe through the study of GWs.

In addition to the so-called ground-based interferometers, a space-based interferometer, LISA (Laser Interferometer Space Antenna), is also planned [70]. The latter, being in space, is not subject to seismic noise, and will therefore be very sensitive to low frequencies in a way that no terrestrial interferometer can currently hope for.

3.1.1 Ground Based Interferometers and The Einstein Telescope

As GWs pass through, they alter the underlying space-time metric, but they do not change the motion of a test particle. To detect and study GWs, the relative motion of two test particles is studied, specifically their geodesic deviation. Although the coordinate distance does not change, the proper distance between two test particles changes proportionally to $h(x, t)$.

Ground-based GW detectors are essentially very large and highly sensitive Michelson interferometers plus a Fabry-Pérot cavity. Currently, five operational GW detectors exist worldwide: two LIGO detectors in the United States [71], Virgo in Italy [72], GEO600 in Germany [73] and KAGRA in Japan [74].

These interferometers use the interference generated by the GW-induced variation of the optical path length with respect to the equilibrium condition. A scheme of a ground-based interferometer is shown in Fig. 3.1. The input laser beam meets a beam-splitter, creating two orthogonal beams. These two beams enter their respective resonant chambers, consisting of a semi-reflective mirror and a reflective one at the end, which is nothing more than the free-falling test mass¹. The signal then returns to the beam-splitter and the twin beams recombine.

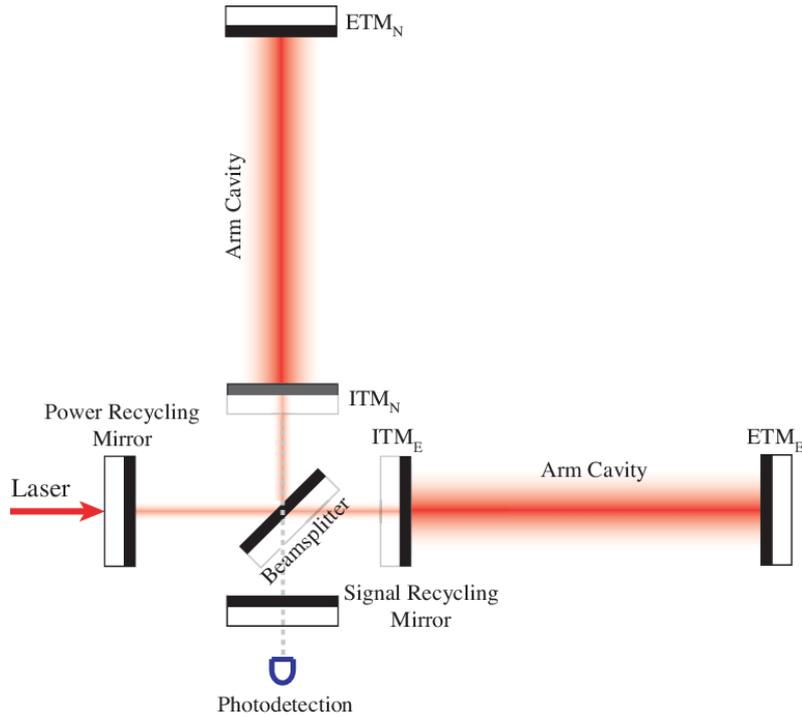


Figure 3.1. A simplified scheme for a ground-based interferometer. Picture from [75].

It is possible to define k_l and ω_l to be the wave number and the pulsation of the laser, respectively. The GW propagates in the z direction and the detector lies in

¹The masses are in quasi-free fall in the direction of propagation of the laser beam, acting as test masses.

the xy plane with the two arms directed as the x and y axes, with the beam splitter at the origin. A photon that travel along the x arm will reach the photodetector after a round trip that that takes $t_0^x + 2L_x/c$ and the same is true for the other arm, i.e. $t_0^y + 2L_y/c$ with $t_0^x \neq t_0^y$, where L_x and L_y are the length of the arms, and t_0 is the time of arrival of the GW. After a round trip, the total electric field is the sum of the two electric fields from each arm, resulting in a power measured by the photodetector, with each field described by eq. (3.1).

$$\begin{aligned} E_1 &= -\frac{E_0}{2} e^{-i\omega_l t + 2ik_l L_x} \\ E_2 &= \frac{E_0}{2} e^{-i\omega_l t + 2ik_l L_y} \end{aligned} \quad (3.1)$$

The cavity amplifies the effect of the incoming GW by making the optical path longer while keeping the detector compact.

Since the initial time t_0 will always be unknown, all the equations are usually expressed in the more physical variable t , that is the time of arrival of the photon. We now focus on a single GW polarisation, $h(t) = h_+(t) = h_0 \cos(\omega_{gw} t)$, the space-time now becomes

$$ds^2 = -c^2 dt^2 + (1 + h_+(t)) dx^2 + [1 - h_+(t)] dy^2 + dz^2. \quad (3.2)$$

At first order in h_0 , we have that for photons travelling along the x axis:

$$ds^2 = 0 \rightarrow dx = \pm c dt \left[1 - \frac{1}{2} h(t) \right], \quad (3.3)$$

where the plus sign is for the first half of the round trip. We need to integrate this equation to obtain the time interval $t - t_0$ for the x -arm, implement a similar integral for the y -arm, and combine these equations to eliminate t_0 .

In the TT gauge, it is possible to show that, at first order in h_0 , the final signal in the photodetector is:

$$E_f(t) = -i E_0 e^{i\omega_l(t-2L/c)} \sin[\phi_0 + \Delta\phi_x(t)], \quad (3.4)$$

where $L = (L_x + L_y)/2$, and $\phi_0 = k_l(L_x - L_y)$ is a constant that can be tweaked, usually to 0. All the signal is therefore in the phase shift $\Delta\phi_x(t)$ that is proportional to

$$\Delta\phi_x(t) \sim h(t - L/c) k_l L. \quad (3.5)$$

By comparing the above formula with the definition of ϕ_0 , the effective signal is proportional to

$$\frac{\Delta(L_x - L_y)}{L} \sim h(t - L/c). \quad (3.6)$$

In the above description, the contribution of the Fabry-Pérot cavity is neglected. The cavity is used to amplify the effect of the incoming GW by "folding" the arm length and thus making the optical path longer, while keeping the detector compact.

So far, only second-generation interferometers with a so-called L-shaped configu-

ration have been discussed. The third generation of detectors is now being designed. The two ground based new detectors, Cosmic Explorer (CE) and ET, will lead the way in not just astrophysical, but cosmological research based on gravitational signals. As can be seen from Fig. 3.2, the third generation is characterised by significantly higher sensitivity.

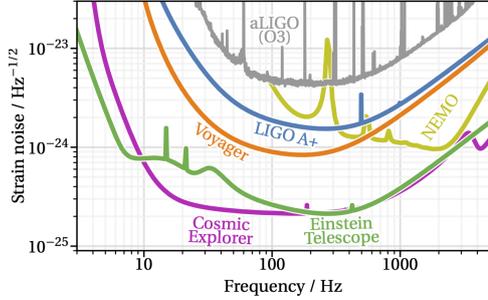


Figure 3.2. Comparison of the sensitivity curves for various GW detectors. The one for Voyager, ET and CE are expected sensitivity curves and not measured. Picture from cosmicexplorer.org/sensitivity.html.

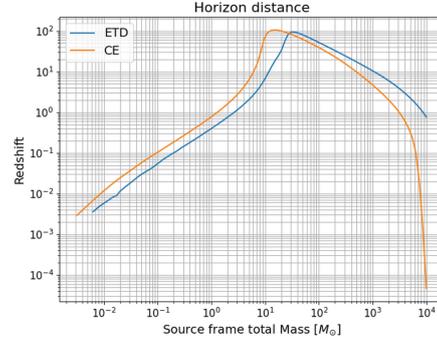


Figure 3.3. The horizons in the equal mass scenario for CE and ET. On the vertical axis is the maximum redshift at which a pair of equal masses BHs can be detected. Picture from [76]

Einstein Telescope Design

The first peculiar feature of ET is its triple detector design, as shown in Fig. 3.4. This innovative setup involves three nested interferometers arranged in an equilateral triangle, each with arms extending several kilometres in length. The three interferometers work as a network, providing a trilateral perspective that enhances the detector ability to triangulate and precisely locate the sources of gravitational waves in the sky. Furthermore, ET will be built in the so-called "xylophone" configuration, meaning each arm will be composed of two cavities, one for low and the other for high frequencies. This configuration enables ET to achieve unprecedented directional sensitivity, allowing astronomers to pinpoint the origin of gravitational wave signals with unparalleled accuracy.

For this type of detector, most of the noise at low frequencies originates from seismic noise, and at high frequencies from thermal noise from the mirrors as well as quantum noise due to the electric field in the photodetector. To mitigate noise as much as possible, ET is planned to be located underground, a strategic choice aimed at reducing the impact of environmental noise and seismic disturbances that can compromise the sensitivity of surface-based detectors. The subterranean placement ensures an exceptionally stable environment, shielding the detector from the myriad of terrestrial vibrations that could interfere with the detection of faint signals. This subterranean infrastructure is aimed at achieving optimal conditions at low frequencies for ET.

To further enhance sensitivity, ET incorporates cryogenic systems designed to cool its detector components to extremely low temperatures. The application

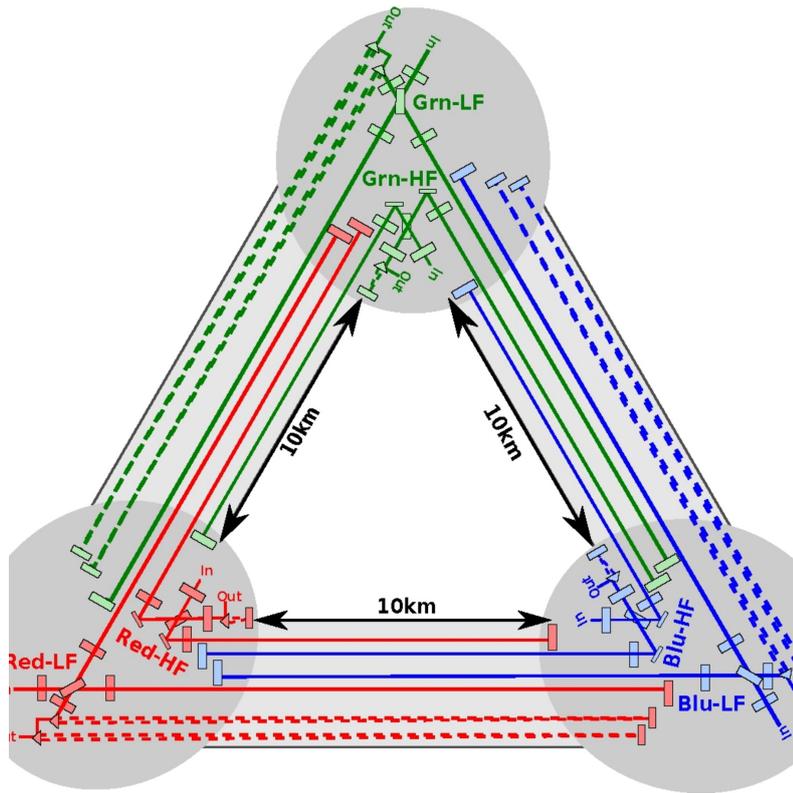


Figure 3.4. A schematic representation of the ET detector in the xylophone configuration. Each arm will be 10 Km long. Picture from [77].

of cryogenics is pivotal in minimising thermal noise, a fundamental limitation in gravitational wave detectors. By reducing the thermal vibrations of the detector components, this technological advancement pushes the boundaries of observational precision, enabling the detection of events that were previously beyond the reach of gravitational wave astronomy [33, 76] as Fig. 3.3 shows.

ET's capability to detect gravitational waves across a broad frequency range opens avenues for cosmological exploration. The telescope can probe the early moments of the universe, offering a unique window into the cosmic dawn, the period from about 50 million years to one billion years after the Big Bang when the first stars, black holes, and galaxies in the Universe formed. Additionally, ET is well-suited to investigate cosmic strings [78], hypothetical one-dimensional structures that could have formed in the early universe. Furthermore, the telescope has the potential to probe the Universe in its early stages, as shown in Fig. 3.5, which will lead to a valuable contribution to cosmological studies.

3.1.2 Antenna Pattern

So far, only the simple case of a GW with a single polarisation and a wave vector perpendicular to the detector plane has been discussed. This simplified scenario was useful to understand the principles behind direct GWs detection.

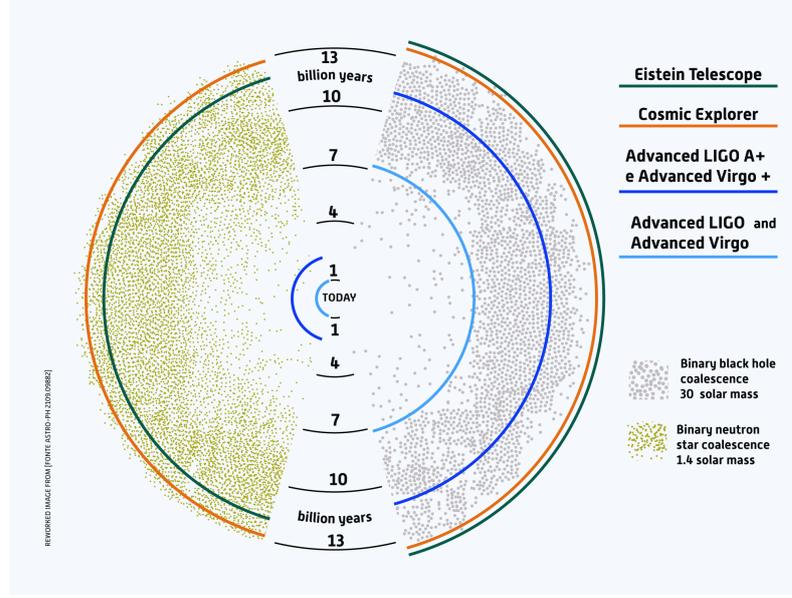


Figure 3.5. A comparison of the distances horizons for actual and future GWs detectors. Picture from einstein-telescope.it/en/einstein-telescope-en/.

In general, a GW will have a wave vector with a random orientation with respect to the detector. Also, considering the symmetries of the system, it is clear that given the detector plane, a wave hitting at an angle φ or at an angle $\varphi + \pi$ will lead to the same $\Delta(L_x - L_y)/L$. It is also worth noting that all planar detectors will have both a preferred and a disfavoured direction of detection: the former corresponds to a wave vector k_{gw} perpendicular to the detector plane, the latter correspond to the case of k_{gw} lying along the detector plane, since GWs are transversal waves.

This behaviour needs to be taken into account, since one important aspect of the signal is its overall amplitude. To take care of the detector response, a function called *antenna pattern* needs to be introduced. The antenna pattern will be first computed for one L-shaped interferometer and then extended to the ET case.

Given a passing GW, one can claim a detection if the signal-to-noise-ratio (SNR) is greater than a threshold SNR_{thr} that depends on the specific detector and a false alarm rate. The SNR is computed in the frequency domain and can be obtained using the following equation:

$$SNR^2 = 4 \int_0^\infty \frac{|\tilde{h}(f)|^2}{S_n(f)} df, \quad (3.7)$$

with $S_n(f)$ representing the detector noise spectra and $\tilde{h}(f)$ denoting the Fourier transformed $h(t)$ tensor multiplied by the antenna function. This function, for a ground-based interferometer, represents the response to each polarisation of the wave. The antenna function depends on the direction of the incoming signal relative to the detector and on the angle ψ , which indicates how much the source reference frame needs to be rotated in order to align the $+$, \times components with the detector

frame, as illustrated in Fig. 3.6. Since the two polarisation states do not mix, \tilde{h} can

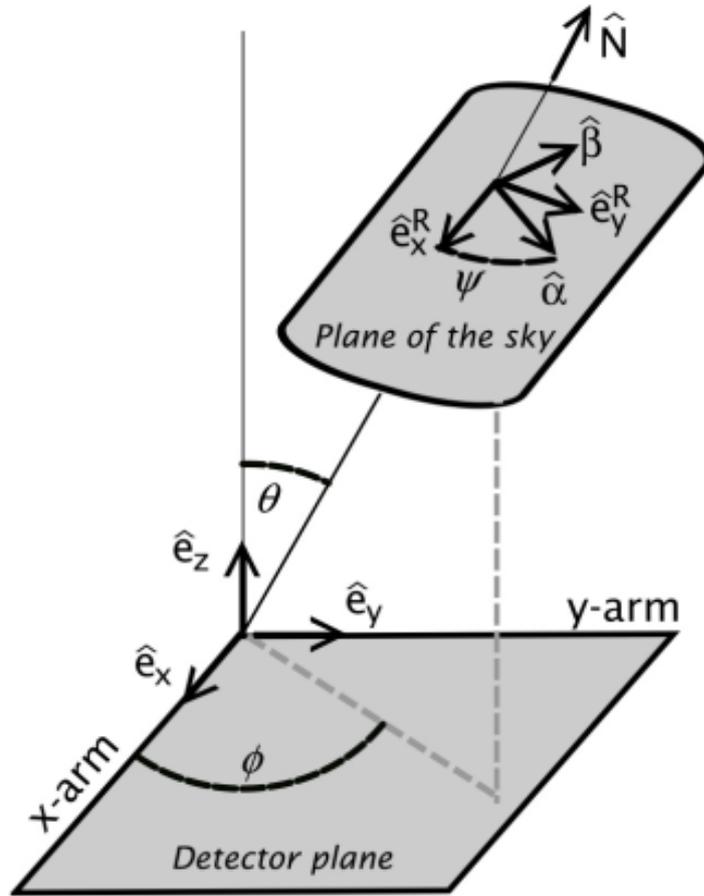


Figure 3.6. The relative orientation between the detector and the approaching GW. Here, ϑ and φ pinpoint the sky position of the GW.

be expressed as,

$$\tilde{h}(f) = F_+(\vartheta, \varphi, \psi)\tilde{h}_+(f) + F_\times(\vartheta, \varphi, \psi)\tilde{h}_\times(f), \quad (3.8)$$

where

$$\begin{aligned} F_+(\vartheta, \varphi, \psi) &= \frac{1}{2} \left(1 + \cos^2 \vartheta \right) \cos 2\varphi \cos 2\psi - \cos \vartheta \sin 2\varphi \sin 2\psi, \\ F_\times(\vartheta, \varphi, \psi) &= \frac{1}{2} \left(1 + \cos^2 \vartheta \right) \cos 2\varphi \sin 2\psi + \cos \vartheta \sin 2\varphi \cos 2\psi. \end{aligned} \quad (3.9)$$

The equations for F_+ and F_\times represent the antenna functions. By averaging over ψ , they form a 3D surface of the antenna pattern.

It is evident that both F_+ and F_\times are never larger than one, thus generally reducing the amplitude of h . Once again, the previously discussed quadrupole formula is used,

expressed in the frequency domain as:

$$\begin{aligned}\tilde{h}_+(f) &= \left(\frac{5}{24}\right)^{1/2} \frac{\pi^{-2/3}}{d_l} \mathcal{M}_c^{5/6} f^{-7/6} \left(\frac{1 + \cos^2 \iota}{2}\right)^2 e^{i\phi(f)} \\ \tilde{h}_\times(f) &= \left(\frac{5}{24}\right)^{1/2} \frac{\pi^{-2/3}}{d_l} \mathcal{M}_c^{5/6} f^{-7/6} \cos^2 \iota e^{i\phi(f) + i\pi/2}.\end{aligned}\quad (3.10)$$

The GW's phase $\phi(f)$ can be neglected, since the relevant quantity is the modulus of h . Here, ι represents the vector normal to the orbital plane of the two inspiraling compact bodies. As in nature it is supposed there not to exist a preferred ι , an average over ι is usually performed in all applications of the above equations. Given that the two GW polarisation states form an orthogonal basis: $F_{+, \times} h_{\times, +} = 0$. Therefore, the SNR integral can be written in a simplified way and, after performing an average over ψ , one obtains:

$$\begin{aligned}\langle SNR^2 \rangle &= 2 \langle F_+^2 + F_\times^2 \rangle \int_0^\infty \frac{|\tilde{h}_+|^2 + |\tilde{h}_\times|^2}{S_n} df; \\ \langle F_+^2 + F_\times^2 \rangle &= P(\vartheta, \varphi) = \frac{1}{4} \left(1 + \cos^2 \vartheta^2\right) \cos^2 2\varphi + \cos^2 \vartheta \sin^2 2\varphi.\end{aligned}\quad (3.11)$$

Since it is clear we are working in the frequency domain, in order to derive eq. (3.11) explicitly from eq. (3.7), the $\tilde{}$ will be removed from \tilde{h} and kept only when necessary. Given the above considerations, the SNR can be rewritten as:

$$\begin{aligned}SNR^2 &= 4 \int_0^\infty \frac{|h(f)|^2}{S_n(f)} df = 4 \int_0^\infty \frac{|F_+ h_+ + F_\times h_\times|^2}{S_n} df = \\ &= 4 \int_0^\infty \frac{F_+^2 |h_+|^2 + \cancel{F_+ h_+ F_\times h_\times^*} + \cancel{F_+ h_+^* F_\times h_\times} + F_\times^2 |h_\times|^2}{S_n} df = \\ &= 4 \int_0^\infty \frac{F_+^2 |h_+|^2 + F_\times^2 |h_\times|^2 + \overbrace{F_\times^2 |h_+|^2}^0 + \overbrace{F_+^2 |h_\times|^2}^0}{S_n} df = \\ &= 4 \left(F_+^2 + F_\times^2\right) \int_0^\infty \frac{|h_+|^2 + |h_\times|^2}{S_n} df.\end{aligned}\quad (3.12)$$

These mathematical steps will always be true for every detector. The mean over the

angular variable ψ is then:

$$\begin{aligned}
\langle SNR^2 \rangle &= \left\langle \frac{4}{2\pi} (F_+^2 + F_\times^2) \int_0^\infty \frac{|h_+|^2 + |h_\times|^2}{S_n} df \right\rangle = \\
&= \frac{4}{2\pi} \int_0^{2\pi} \int_0^\infty (F_+^2 + F_\times^2) \frac{|h_+|^2 + |h_\times|^2}{S_n} df d\psi = \\
&= \overbrace{4 \int_0^\infty \frac{|h_+|^2 + |h_\times|^2}{S_n} df}^\xi \frac{1}{2\pi} \int_0^{2\pi} (F_+^2 + F_\times^2) d\psi = \\
&= \frac{\xi}{2\pi} \left(\int_0^{2\pi} F_+^2 d\psi + \int_0^{2\pi} F_\times^2 d\psi \right) = \tag{3.13} \\
&= \frac{\xi}{2\pi} \left(\int_0^{2\pi} (A^2 \cos^2 2\psi + B^2 \sin^2 2\psi + AB \sin 2\psi \cos 2\psi) d\psi + \right. \\
&\quad \left. + \int_0^{2\pi} (A^2 \sin^2 2\psi + B^2 \cos^2 2\psi + AB \sin 2\psi \cos 2\psi) d\psi \right) = \\
&= \frac{\xi}{2} (A^2 + B^2).
\end{aligned}$$

This result indeed matches eq. (3.11), after substituting A , B , and ξ . This result holds whenever one is not interested in the relative orientations between the detector and the GW source. A similar logic can be applied for the average over the other two angular variables θ, ϕ .

A Network of Detectors

We now apply the technique shown before to a network of GWs detectors. The concept is that ET is essentially a network of three detectors, so a method is needed to combine multiple interferometers. Since there is no formal difference in writing the equations for a network of detectors placed in different locations (i.e., LIGO(H), LIGO(L), Virgo, and KAGRA) or concentrated in a small area, like ET will probably be, a general mathematical formulation will be adopted, finally applying the result to the ET specifics.

As stated in [79], a common system of reference (RS) is required. This SR's origin is usually set at the centre of the Earth, and latitude and longitude are used to pinpoint each interferometer. The main goal here is to describe the antenna pattern of each detector using this new RS.

Table 3.1.

The values of χ and position of each detector in the network in degrees.

	$\chi(^{\circ})$	$\beta(^{\circ})$	$\lambda(^{\circ})$
Ligo H	171.8	46.45	-119.41
Ligo L	243	30.56	-90.77
Virgo	116.5	43.63	10.5
Kagra	70	36.25	137.18

Under the above assumptions, the resulting expression for the antenna pattern of an arbitrarily located and oriented L-shaped interferometer is as follows. The source position is given by the spherical coordinates ϑ, φ on the sky, and the frame for the wave polarisation angle ψ is defined to be aligned with this spherical-coordinate grid. The detector is at latitude β and longitude λ . The interferometer is oriented so that the bisector of its arms points in the direction χ , measured counter-clockwise from East. Its arms have an opening angle of η . The celestial coordinates ϑ, φ are aligned with latitude and longitude, so that the equators of both systems coincide and the celestial point $\vartheta = \pi/2, \varphi = 0$ is in the zenith direction above the geographic location $\beta = 0, \lambda = 0$. We will set $\eta = \pi/2$ since L-shaped interferometers satisfy this condition. The antenna pattern functions now become:

$$F_{+} = \sin(\eta) [a \cos(2\psi) + b \sin(2\psi)] \quad (3.14)$$

$$F_{\times} = \sin(\eta) [b \cos(2\psi) - a \sin(2\psi)] , \quad (3.15)$$

with a and b defined by

$$\begin{aligned} a = & \frac{1}{16} \sin(2\chi) [3 - \cos(2\beta)] [3 - \cos(2\vartheta)] \cos[2(\varphi + \lambda)] \\ & + \frac{1}{4} \cos(2\chi) \sin(\beta) [3 - \cos(2\vartheta)] \sin[2(\varphi + \lambda)] \\ & + \frac{1}{4} \sin(2\chi) \sin(2\beta) \sin(2\vartheta) \cos(\varphi + \lambda) \\ & + \frac{1}{2} \cos(2\chi) \cos(\beta) \sin(2\vartheta) \sin(\varphi + \lambda) + \frac{3}{4} \sin(2\chi) \cos^2(\beta) \sin^2(\vartheta) \end{aligned} \quad (3.16)$$

$$\begin{aligned} b = & \cos(2\chi) \sin(\beta) \cos[2(\varphi + \lambda)] \\ & - \frac{1}{4} \sin(2\chi) [3 - \cos(2\beta)] \cos(\vartheta) \sin[2(\varphi + \lambda)] \\ & + \cos(2\chi) \cos(\beta) \sin(\varphi + \lambda) - \frac{1}{2} \sin(2\chi) \sin(2\beta) \sin(\vartheta) \sin(\varphi + \lambda) . \end{aligned} \quad (3.17)$$

The coordinates of each detector in the network are needed, which can be found in [80].

Entering the data in Tab. 3.1 into these new set of equations, the SNR in each detector can be obtained. For ET, one has to consider each detector as a separated unit, then set $\eta = \pi/3$ and use the right values of β and λ . Since the three detectors will probably be at the same location, one can simplify all the above equations and,

using eq. (3.7), obtain the ET antenna function as

$$\begin{aligned}
F_+^{(1)}(\vartheta, \varphi, \psi) &= \frac{\sqrt{3}}{2} \left[\frac{1}{2} \left(1 + \cos^2 \vartheta \right) \cos 2\varphi \cos 2\psi - \cos \vartheta \sin 2\varphi \sin 2\psi \right], \\
F_\times^{(1)}(\vartheta, \varphi, \psi) &= \frac{\sqrt{3}}{2} \left[\frac{1}{2} \left(1 + \cos^2 \vartheta \right) \cos 2\varphi \sin 2\psi + \cos \vartheta \sin 2\varphi \cos 2\psi \right], \\
F_{+,\times}^{(2)}(\vartheta, \varphi, \psi) &= F_{+,\times}^{(1)}(\vartheta, \varphi + 2\pi/3, \psi), \\
F_{+,\times}^{(3)}(\vartheta, \varphi, \psi) &= F_{+,\times}^{(1)}(\vartheta, \varphi + 4\pi/3, \psi).
\end{aligned} \tag{3.18}$$

This is essentially the antenna pattern for an L-shaped interferometer with an overall prefactor $\sin(\eta) = \sin(\pi/3)$. The superscripts refer to the individual antenna.

Lastly, there is no correlation between the noise in the network, the SNR for a collection of N interferometers is defined by

$$\rho_{tot} = \sqrt{\sum_1^N \rho_i^2}, \tag{3.19}$$

where ρ_i^2 is the square of the SNR in the i -th detector.

Events Properties

In this section, I present some typical parameters for DS events, primarily based on [81]. The focus is on three major quantities: the sky localisation accuracy, the luminosity distance uncertainty, and the number of detected sources per year. It is assumed that the detector has a triangular shape in the xylophone configuration.

The most interesting events are the so-called "golden events". A binary black hole (BBH) event is a golden-event if it has one of the following characteristics: SNR equal to or larger than 100, a sky localisation area $\Omega_{90} \leq 10 \text{ deg}^2$ (meaning the 90% contour probability area is less or equal to 10 deg^2), or a relative uncertainty for the luminosity distance $\Delta dl/dl \leq 0.05$. Typically, these characteristics are correlated, and if an event meets one requirement, it usually satisfies the others. For the purpose of this thesis, the standard event has $\Delta dl/dl = 0.1$ and $\Omega_{90} = 10 \text{ deg}^2$, so the focus is not on the best case scenario. Only after showing that the posterior is dominated by golden events, the uncertainty on the luminosity distance will be set to $\Delta dl/dl \leq 0.05$. The next chapter will compare this standard case with more optimistic and pessimistic cases.

Given the small area covered by Ω_{90} , it is assumed that the sky map follows a two-dimensional Gaussian distribution over the angular variables, centred at the DS position. The distribution is constructed such that $\Omega_{90} = 10 \text{ deg}^2$.

Figure 10 of [81], included here, shows the redshift distribution of golden-events per year. It is concluded that about $\mathcal{O}(200)$ - $\mathcal{O}(400)$ golden-events are expected per year, given the redshift range of the Euclid flagship.

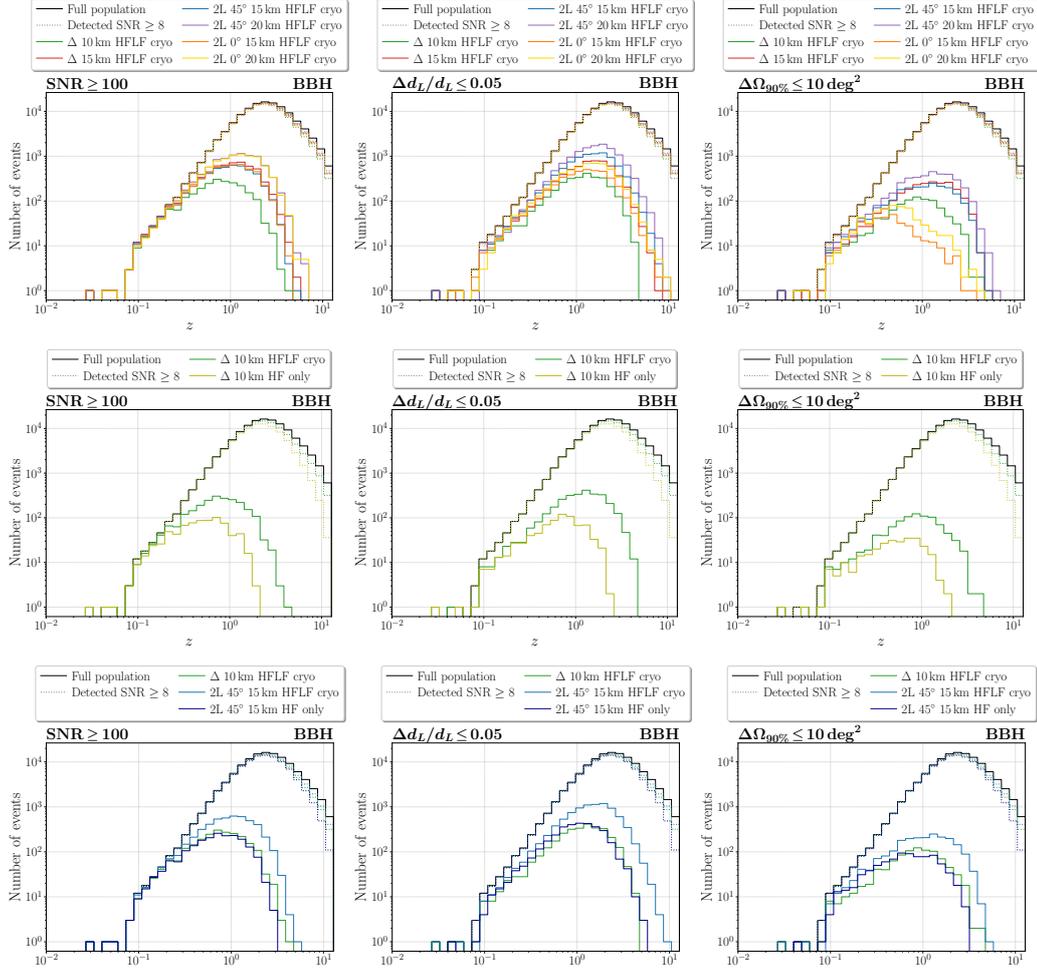


Figure 3.7. Redshift distribution of golden-events. The left column shows the distribution of BBHs merges with $\text{SNR} \leq 100$. The central column shows the distribution for $\Delta dl/dL \leq 0.05$ and the right column for $\Omega_{90} \leq 10 \text{ deg}^2$. Picture from [81].

3.2 Euclid

Euclid is a medium-class (M-class) mission and is part of the Cosmic Vision campaign of European Space Agency (ESA) Science Programme. It represents a new generation of space telescopes, designed to study both dark energy and dark matter. Launched on July 1st 2023, the Euclid mission aims at enhancing our understanding of the fundamental constituents shaping the Cosmos.

The central focus of Euclid is the investigation of the recent accelerated expansion of the universe, nowadays attributed to the presence of dark energy or to a modification of gravity. Employing a combination of imaging and near-infrared photometry and spectroscopy, the mission seeks to map the distribution of galaxies across the cosmic time. By scrutinising the three-dimensional structure of the universe on a large scale, Euclid aspires to trace the evolution of cosmic structures and shed light on the underlying physics of dark energy.

To achieve its scientific objectives, the Euclid mission employs a suite of advanced instruments and observational methodologies:

- **Imaging and Photometric surveys:** Euclid’s space telescope is equipped with a state-of-the-art visible camera, allowing for precise imaging of galaxies. This capability enables the measurement of the shapes and sizes of galaxies across different redshifts, providing crucial information for understanding their evolution. The imaging Euclid catalogue will contain the shapes of about 1.5 billion galaxies, observed in the visible range with the Visual Imaging (VIS) instrument, while the redshifts of such galaxies will be measured in the photometric mode, using the Near Infrared Spectroscopic Photometric (NISP-P) instrument, complemented by ground-based observations in different bands
- **Spectroscopic Redshift Survey:** The mission incorporates spectroscopic observations via the NISP-S instrument, to determine the redshifts of galaxies accurately. This information is essential for constructing a three-dimensional map of the cosmic structure, allowing scientists to trace the expansion history of the universe. The principal spectral characteristic employed for redshift determination is the H α line. This originates from the transition between $n = 3$ and $n = 2$ energy levels in hydrogen, exhibiting a wavelength of approximately $\lambda \simeq 656.3$ nm. Specifically, the scientific requirement concerning the galaxy number density within the spectroscopic sample necessitates that the average effective H α flux limit for a source with a 1 arcsec diameter should not exceed $3e - 16 \text{ erg cm}^{-2} \text{ s}^{-1}$ at 1600 nm. The flux threshold² encapsulates the line flux above which galaxies are reliably detected. Owing to the abundance and the dimensions of the spectra discernible within the field of view, numerous spectra are likely to be contaminated—or confused—by those from adjacent galaxies. Such contamination precipitates inaccuracies in redshift assessments, implying that not all observed spectra will facilitate a trustworthy extraction of redshift data.
- **Survey Strategy:** Euclid employs a systematic survey strategy, covering a large portion of the sky to ensure statistical robustness. The mission’s wide-field survey, combined with its depth and accuracy, enables comprehensive studies of dark energy and cosmic structure evolution over cosmic time.

3.2.1 FlagShip

The possibility to cross correlate with a spectroscopic set of galaxies, and the accuracy of the 3D map of the hosts, is a crucial feature for this work. More important, Euclid will cover about 15000 deg² of the sky area, making this survey an outstanding host catalogue. More can be found in the ESA web page about Euclid or the Euclid red book

²The flux threshold denotes the minimal line flux required for galaxies to be detectable, ensuring an SNR greater than 3.5.

The telescope is now launched, and real data will arrive in 2024. For this work, a mock catalogue of galaxy hosts, the Euclid flagship, has been used. The spectroscopic flagship is a collection of galaxies with spectroscopic measurements of z . This dataset is a simulation of what the satellite will see and is the best spectroscopic "observed" collection of galaxies one can use to predict cross correlation on the Euclid footprint.

Developed by Euclid's teams, the groundbreaking Euclid Flagship mock galaxy catalogue [82] is based on a supercomputer N-body simulation ran in a box of $3600 h^{-1}$ Mpc on a side with 16000^3 particles each one with a mass of $10^9 h^{-1} M_{\odot}$, allowing one to resolve dark matter halos with a minimum mass of $10^{11} h^{-1} M_{\odot}$, matching the typical size of the halos capable of hosting the faintest galaxies expected to be detected in the Euclid survey.

The simulation assumes a reference cosmological with parameters similar to those estimated by the Planck collaboration [17] namely:

- $\Omega_m = 0.319$
- $\Omega_b = 0.049$
- $\Omega_r = 0.00005509$
- $\Omega_{\nu} = 0.00140343$
- $\Omega_{\Lambda} = 1 - \Omega_m - \Omega_r - \Omega_{\nu}$
- $T_{cmb} = 2.7255$ K
- $H_0 = 67$ Km/s/Mpc
- $n_s = 0.96$
- $\sigma_8 = 0.813$
- $\omega_{de} = -1.0$

The initial conditions were set at $z = 99$ by applying first-order Lagrangian perturbation theory displacements from a uniform particle grid.

The outputs of the simulations were arranged to a full-sky particle light-cone (to $z = 3$). The full-sky coverage was achieved by replicating the simulation box. The resulting light cone contains 31 trillion particle with specified positions and peculiar velocities. The ROCKSTAR [83] halo finder was then used to extract the ~ 150 billion haloes that contain the mock galaxies. In this Thesis, I will use a smaller catalogue covering one octant of the sky, since this is the reference catalogue currently validate by the Euclid collaboration and routinely used to validate the data analysis pipeline.

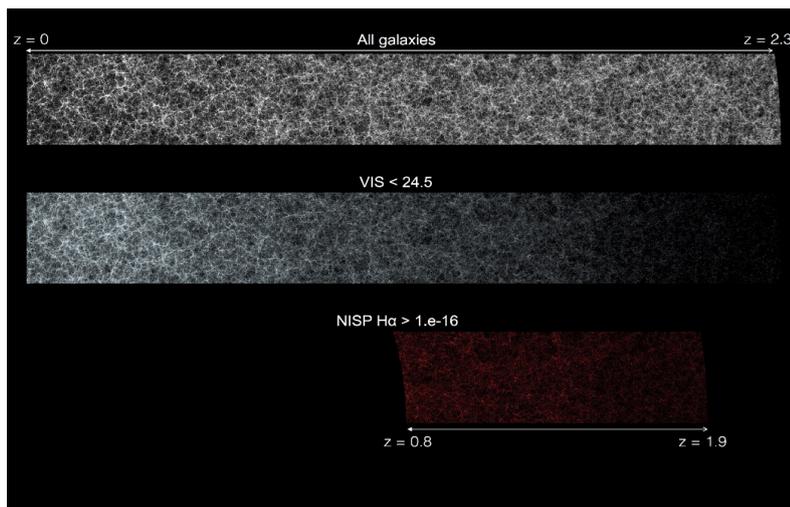
Once the halo catalogue is generated and its validity checked by comparing the halo mass function to theoretical expectations [84], using a prescription that combines the halo occupation distribution model [85] and abundance matching techniques, see [86] and references therein. This hybrid method computes the number of satellites in each halo and assigns the luminosities to central and satellite galaxies. Galaxy

clustering measurements are used to determine the free parameters and relations implemented. Galaxy colours are then assigned using the colour-magnitude relation, similarly to what was done in [87], taking into account whether the galaxy is central or satellite [88]. While the central galaxies are always placed at the centre of the halo, satellite galaxies are located following a triaxial Navarro, Frenk, and White profile [89]. Galaxy velocities also depend on their type. Central galaxies are assumed to have the velocity of the halo centre of mass. The velocities of the satellite galaxies are estimated by solving the spherical, stationary, Jeans equation of local dynamical equilibrium. Galaxy velocities are used to compute the observed redshifts listed in the mock catalogue.

The spectral energy distribution (SED) of the mock galaxies were obtained from two template libraries: COSMOS SED [90] and the Bruzual & Charlot one [91]. The result is 136 templates obtained using the 31 SEDs and the two Galactic extinction laws. The SED is assigned to each galaxy depends on the redshift, and the colour of the object. Since the Euclid spectroscopic survey preferentially target H α line emitting galaxies, additional physical parameters need to be assigned to the mock galaxies. The first step is to specify the ultraviolet flux that is then used to estimate the star formation rate and the stellar mass.

The flux of the H α line is estimated from the star formation rate using the Kennicutt law [92]. The flux of the other emission lines is also estimated after computing the galaxy metallicity from the galaxy stellar mass, following the relation obtained in [93].

The result is a catalogue of mock galaxies with an H α flux above 2×10^{-16} ergs $^{-1}$ cm $^{-2}$ in the redshift range, $z = [0.9, 1.8]$ matching the expected properties of the Euclid spectroscopic surveys. This is the mock sample used in my Thesis. To conclude this section, I leave a picture of the mock catalogue, with the spectroscopic sample highlighted in red.



Credits P. Fosalba

Figure 3.8. A representation of the flagship simulation. The spectroscopic sample is highlighted in red.

Chapter 4

The Hierarchical Bayesian Formalism

The primary aim of this chapter is to demonstrate how to estimate the H_0 via the cross-correlation technique. The discussion here focuses exclusively on astronomical sources of GWs, particularly those arising during the inspiral and coalescence of two compact objects. The three main event categories include BHBH, BHNS, and NSNS.

The goal is to determine both the redshift z and the luminosity distance of the merging. GWs events fall into two categories: those with an electromagnetic counterpart, namely BS, and those without, known as DS. BS are critical for accurately establishing the $z - Dl$ relationship and determining the parameters of the underlying cosmological model.

The method involves analysing the electromagnetic spectrum to identify known emission or absorption lines in the received light. Redshift is then calculated by comparing the measured line wavelength with the known one. The method relies on gathering sufficient BS events in order to constrain cosmological parameters. The precision of this method is limited only by the number of observed BS and the accuracy in determining both Dl and z . This approach was applied to the significant event *GW170817* [4] in August 2017. This event marked the first of its kind, offering an opportunity to study the tidal deformability of neutron stars, a parameter intimately connected to the behaviour of matter under extreme conditions. This single event allows an H_0 estimation and its comparison with measurements from other cosmological observable. However, as illustrated in Fig. 4.1, the posterior from this single even is not enough to resolve the Hubble tension. The 68.3% credible interval for the posterior yields a Hubble constant of

$$H_0 = 70_{-8.0}^{+12.0}, \text{ km, s}^{-1}, \text{ Mpc}^{-1},$$

aligning broadly with both previously mentioned estimations of H_0 .

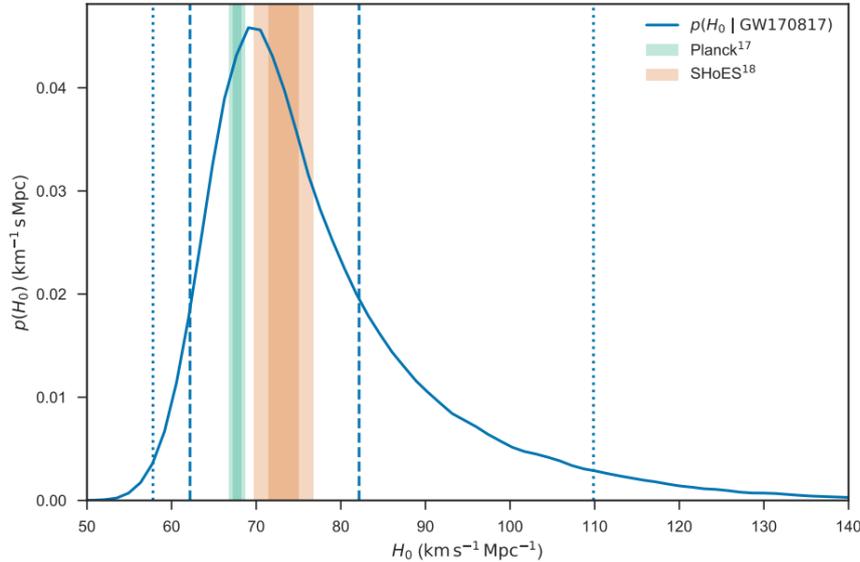


Figure 4.1. The posterior distribution (blue curve) for H_0 obtained from *GW170817* by the LIGO-Virgo collaboration, compared with two other measurements of the same parameter from Planck and SHoEs collaboration. Figure from [94]

4.1 A New Tool

While NSNS events are promising, they are rare as compared to BHBH events [95]. Considering potential light emissions from BHNS events, relying on a sufficient number of mergers is not feasible for this type of research. In contrast, dark events are approximately 100 times more frequent and offer a practical alternative to H_0 measurements. DS primarily involve coalescence events of two astrophysical black holes or NSBH/NSNS events too distant for their electromagnetic counterpart to be visible on Earth. This work will focus solely on BHBH events.

As mentioned in the previous chapter, the spiral motion of two bodies towards each other generates a distinct gravitational wave signal. The luminosity distance can be estimated as previously mentioned, but in the absence of electromagnetic radiation, redshift information is missing. The problem is that both a loud source with a large chirp-mass and a faint one can produce signals of similar amplitude when redshift effects are included. To address this, one can either use information about the redshift distribution of BH masses i.e. Spectral Sirens [96, 97] or a non-homogeneous distribution of sources. In the following I will explain how, based on the knowledge of the source distribution, the degeneracy can be broken to obtain redshift information.

Black holes represent the endpoint of stellar evolution, and given the non-uniform star formation rate (SFR) as a function of z , a peak in the astrophysical black hole distribution is expected. Assuming the SFR follows the model in [98],

$$\psi(z) = 0.015 \frac{(1+z)^{2.7}}{1 + \left(\frac{1+z}{2.9}\right)^{5.6}}, \text{M}_{\odot} \text{Mpc}^{-3} \text{yr}^{-1}, \quad (4.1)$$

it becomes necessary to account for a time delay between star formation and merging. This means waiting for stars to collapse into black holes. The distribution of this time delay, $\mathcal{P}(t)$, remains uncertain, but an effective time can be employed to model this effect. By defining

$$\varphi(z) = \frac{t_H}{(1+z)E(z)}, \quad (4.2)$$

$$t(z) = \int_z^\infty \varphi(x) dx, \quad (4.3)$$

where $t_H = 1/H_0$ is the Hubble time and the second equation represents the cosmic time as a function of redshift, one can write the GW rate, R_{GW} , as a function of time:

$$R_{GW}(t) = \mathcal{N} \int_t^\infty \psi(x) \mathcal{P}(t-x) dx. \quad (4.4)$$

The physical meaning of R_{GW} is that, at each time, one must first compute the total amount of stellar mass available, introduce a delay until the star collapses, and then calculate the amount of radiation receivable at a certain time due to the finite propagation velocity of GWs.

Various prescriptions exist for the distribution $\mathcal{P}(t)$; a reasonable assumption is that the process of decaying into a BH and merging is rare, thus following a decay-like equation as in [99, 100]. Therefore, one can assume

$$\mathcal{P}(t(x) - t(\tilde{x})) = \frac{1}{\tau} e^{-\frac{t(\tilde{x}) - t(x)}{\tau}}, \quad (4.5)$$

where τ is a parameter representing the delay time, ranging from 0.1 Gyr to the present epoch.

One must then define a comoving merger rate \mathcal{R} and the observed rate R_{GW} via

$$R_{GW}(z) = \frac{dN}{dt_d dz} = \frac{1}{1+z} \frac{dV}{dz} \mathcal{R}(z), \quad (4.6)$$

such that, with a slight redefinition of the time $t(z)$ as $\int_0^z \phi(\tilde{z}) d\tilde{z}$, the rate per comoving volume is given by:

$$\mathcal{R}(z) = \frac{1}{\tau} \int_z^\infty \frac{dt}{d\tilde{z}} \psi(\tilde{z}) e^{-\frac{t(\tilde{z}) - t(z)}{\tau}} d\tilde{z}. \quad (4.7)$$

Given the merger rate, it becomes possible to estimate H_0 using Bayesian inference as detailed in [99]. The specifics of this measurement are beyond the scope of this work (for further discussion, see [99, 100, 101, 102]). In the next chapter I show how the particular distribution of sources has not a significant impact, rather what matters is the number of gold events within a given redshift range. The specific value of the delay time becomes less crucial unless it significantly alters the number of detected events.

Before discussing the approach actually used in this thesis, it is worth briefly describing what Spectral Sirens are. Like previously mentioned, they belong to the category of dark sources. As with the case just described, this type of source

also attempts to exploit a particular characteristic, in this instance in the mass distribution of the sources, to try to break the $z - \mathcal{M}c$ degeneracy [96, 97]. The features present in the mass spectrum, usually gaps due to the physics behind the formation of stellar-origin black holes, allow for the anchoring of the mass determination in the source's frame of reference, and subsequently in that of the detector, thereby determining the redshift of the binary system.

As mentioned above, another approach to obtain redshift information involves statistically assigning a host galaxy to the GW event. If the galaxy redshift is known and astrophysical BH are assumed to reside inside or near galaxies, both z and Dl can be recovered, enabling the inference of H_0 . This technique requires a galaxy catalogue to be correlated with the DS position, and is commonly referred to as the *cross-correlation* method. The concept is well-established [6, 103, 7, 104, 8, 105] using hierarchical Bayesian models as outlined in [106].

4.1.1 Hierarchical inference

The hierarchical Bayesian model framework applies whenever the goal is to infer "hyperparameters" from data. A hyperparameter, such as the mass distribution of the black hole population, extends beyond the original parameter set regulating individual events but can be deduced from multiple events.

Formally, the framework can be formalised as

$$P(\Lambda|data) = \frac{\mathcal{L}(data|\Lambda)\Pi(\Lambda)}{\int d\Lambda \mathcal{L}(data|\Lambda)\Pi(\Lambda)}, \quad (4.8)$$

where $\Pi(\Lambda)$ represents the hyper-prior in the hyperparameter Λ , and $P(\Lambda|data)$ denotes the hyper-posterior. Given a prior $\Pi(\theta|\Lambda)$, the likelihood \mathcal{L} can be defined as,

$$\mathcal{L}(data|\Lambda) = \int d\theta \mathcal{L}(data|\theta)\Pi(\theta|\Lambda). \quad (4.9)$$

Here, θ is the set of all parameters relevant to the event.

This approach is detailed in the following paragraph, explaining all the terms involved and providing a general overview. The ultimate goal is to obtain,

$$P(H_0|\{D_{GW}\}) = \frac{\mathcal{L}(\{D_{GW}\}|H_0)\Pi(H_0)}{P(\{D_{GW}\})}. \quad (4.10)$$

Here, $\Pi(H_0)$ is a flat prior for H_0 , ranging from some minimum to maximum value.

The likelihood $\mathcal{L}(\{D_{GW}\}|H_0)$ is a combination of a finite ensemble of uncorrelated GW events

$$\mathcal{L}(\{D_{GW}\}|H_0) = \prod_{i=1}^{N_{obs}} \mathcal{L}(\{D_{GW}^i\}|H_0). \quad (4.11)$$

In principle, the effect of the rate R_{GW} should be included. However, given its negligible contribution to the overall result, as stated in [107, 103], eq. (4.10) is used in this work. It is also worth mentioning that equation (4.11) holds under the assumption that each DS is uncorrelated from all the others, i.e., there are not two

DS sharing a close host. In nature, this naturally occurs due to the large size of the Universe. However, when using mock data with a huge number of DS, this problem arises. As stated in [103], this effect is negligible compared to other sources of error, so eq. (4.11) is assumed to be valid.

$\mathcal{L}(\{D_{GW}^i\}|H_0)$ can be obtained by marginalization, i.e., after adopting a lighter notation, $\mathcal{L}(\{D_{GW}^i\}|\theta, H_0)$ is defined as $\mathcal{L}_i(\theta, H_0)$:

$$\mathcal{L}(\{D_{GW}^i\}|H_0) = \frac{1}{\beta(H_0)} \int d\theta \underbrace{\mathcal{L}_i(\theta, H_0)}_{\text{likelihood}} \overbrace{\Pi(\theta|H_0)}^{\text{prior}}. \quad (4.12)$$

The normalisation β is set aside for the moment. All quantities needed to evaluate $\mathcal{L}_i(\theta, H_0)$ are provided in the GW data. The likelihood \mathcal{L}_i is well described by a Gaussian likelihood over the luminosity distance

$$\mathcal{L}_i(\theta, H_0) = \mathcal{L}_i(z, \Omega, H_0) = \rho_i(\Omega) \frac{\mathcal{N}_i(\Omega)}{\sqrt{2\pi}\sigma_i(\Omega)} e^{-\frac{(Dl(z, H_0) - \mu_i(\Omega))^2}{2\sigma_i^2(\Omega)}}. \quad (4.13)$$

The terms of (4.13) are:

- Dl is the luminosity distance.
- μ_i is the measured luminosity distance for the i -th event.
- σ_i is the error for the i -th event.
- ρ_i is the probability of an event occurring in the Ω direction.
- \mathcal{N}_i normalises so that \mathcal{L}_i is a PDF when marginalised over the whole solid angle.

While in general ρ can be an arbitrary distribution in the angular variables, in this thesis $\rho(\Omega)$ is a multivariate symmetric Gaussian on the celestial vault.¹ the function will assign a weight to each host in accordance with the Gaussian distribution. As is usual in this field, when the location region of an event is given, it is intended to indicate the region of the sky that contains 90 % of the total probability. From this information and assuming equal errors on ϕ and θ , it is possible to infer the σ of the Gaussian distribution on the sphere. This is the method used in Chapter 5, taking care to use values consistent with the accuracy involved. Instead, in this chapter ρ is a Gaussian as described before, but its width is exaggerated in order to check its effect on the final result, as illustrated in Fig. 4.6 in which the possible host are placed on different line of sight, hence the different eight and at different redshift therefore they contribute to different values of H_0 .

With the prior in eq. (4.12), one can evaluate eq. (4.13) for each galaxy in the sky volume determined by the GW signal. The luminosity distance of the GW, that

¹To be precise, one should use a Von-Mises Fisher distribution. It is proven that for large enough concentration parameter, the two distributions are the same and the relation $k \sim 1/\sigma$ holds true.

is a constant in H_0 , and the host redshifts are treated as input data in the following analysis.

As it is evident from eq. (4.13), by varying H_0 , for given μ and z , the likelihood follows a Gaussian-like profile, peaking at some value of H_0 . The likelihood for a single event is then a combination of Gaussian functions, with a peak at the "true" value of H_0 plus other peaks acting as white noise. When more likelihoods are multiplied together as in eq. (4.11), the white noise contribution is reduced, since the peaks will correspond to different values of H_0 . The only common contribution is that due to the true value of H_0 . The higher the number of DS, the more distinct the peak for the true value of H_0 becomes. A more in-depth discussion about the properties of this method will be given later in Sec.4.1.4.

4.1.2 Galaxy Catalogue

From the previous section, it is understood that the prior $\Pi(\theta|H_0)$ is essential for the cross-correlation technique. The parameter vector θ can be factorised, focusing on the $\Pi(\theta|H_0)$ dependency on z and Ω , the latter being the angular part of the galaxy prior. Hence, the prior can be written as

$$\Pi(\theta|H_0) = P_{cat}(z, \Omega)P_{\theta'}(\theta'). \quad (4.14)$$

It is noted that, in the context of a redshift galaxy survey, the galaxy prior, $P_{cat}(z, \Omega)$, does not depend on the Hubble constant H_0 . This factorisation is valid under the assumption that all other parameters in θ' are not correlated with either the redshift or the sky localisation. While this is a robust assumption for the sky localisation, some models might introduce a correlation between, e.g., the masses of the BHs and the redshift, thus potentially making incorrect the factorisation above. This work assumes that the BHs mass distribution does not evolve in redshift.

The prior P_{cat} denotes the probability of locating a galaxy at a specific spot in the sky and at a particular redshift. In an ideal catalogue with N galaxies at infinitely precise positions, the prior is defined as

$$P_{cat}(z, \Omega) = \frac{\sum_{\alpha=1}^N w_{\alpha} \delta(z - z_{\alpha}) \delta(\Omega - \Omega_{\alpha})}{\sum_{\alpha=1}^N w_{\alpha}}, \quad (4.15)$$

where w_{α} enables different weighting for each galaxy, and δ represents the Dirac delta. The weights ω can either be a *statistical* weight, accounting for missing galaxies, or an *intrinsic* weight, reflecting the likelihoods of different galaxies hosting a DS. It is reasonable to suppose that more massive galaxies, with a larger number of stars, are more likely to harbour BHs and thus binary systems of BHs (BBHs). In this study, it is considered that all galaxies have the same intrinsic weight, which serves as a solid starting point. Introducing a variable intrinsic weight would not alter the formalism.

To take into account that no real catalogue has an infinite precision on z and Ω , it is always possible to smear the Dirac deltas in eq. (4.15) with, e.g., a Gaussian distribution centred around the true value. Although this is possible, in this work

we used the Flagship spectroscopic galaxy catalogue, with errors over z that are negligible with respect to the errors of the luminosity distance measured from a DS, so we can keep the Dirac delta over the redshift variable, and the same approximation is even better for the angular variables. To include the smearing, one can just write

$$P_{cat}(z, \Omega) = \frac{\sum_{\alpha=1}^N w_{\alpha} \mathcal{G}(z - z_{\alpha}) \mathcal{G}(\Omega - \Omega_{\alpha})}{\sum_{\alpha=1}^N w_{\alpha}}, \quad (4.16)$$

where \mathcal{G} is, for example, a Gaussian distribution around the true value.

No catalogue can contain every galaxy in the Universe. This implies that some galaxies, potentially hosting a DS, will always be missed. In [103, 6] a catalogue is defined complete if all the DSs have the "true" host in the galaxy catalogue. This concept differs from the notion of completeness in a galaxy survey, which typically measures the light flux up to a minimum threshold F_{min} . A survey is considered flux-complete up to F_t , meaning it includes every object with $F_t \geq F_{min}$.

While this could be a concern, the spatial correlation of galaxies, as noted in [108], suggests that even if P_{cat} does not include the actual host galaxy of a DS, there will likely be a galaxy correlated to it. This degrades the final posterior for an individual event, but with a large number of sources, this mainly results in the need for a larger amount of DS.

Two methods exist for accounting for missing galaxies. The first involves estimating the probability of missing a galaxy, as explored in [6]. Knowing the expected density of galaxies allows for calculating the number of galaxies in a region and, by subtracting the observed number, the probability of missing a galaxy can be inferred. In the following, I adopt a similar approach, based on the assumption that the Universe is uniform.

A galaxy survey essentially consist of angular positions and redshifts. By making a histogram of the distribution of objects in the catalogue as a function of z , one obtains the selection function $N(z)$. Under the hypothesis of a uniform Universe that "covers" the survey, as depicted in Fig. 4.2, this set of galaxies is referred to as the true Universe or uniform Universe, being uniform in comoving volume and denoted by U . Given that $N(z)$ is an observation of $U(z)$, then $U(z) \geq N(z), \forall z$. The statistical weight for each galaxy is then defined as the ratio of the number of observed galaxies to the expected number.

$$\omega(z) = \frac{U(z)}{N(z)}. \quad (4.17)$$

While it might seem that there is a degree of arbitrariness in creating a uniform universe without explicitly fixing its density, the only constraint applied here is that for each redshift, it contains more galaxies than the considered survey. However, this does not pose a problem because the term P_{cat} consistently appears both in the numerator and in the normalisation factor β in eq. (4.12). This cancels out the dependence on the absolute number of objects in the catalogue. To facilitate faster computation of the final posterior, the uniform Universe is constructed with the same density as the densest z -bin of the observed Universe.

The physical significance of the statistical weight is that it allows each galaxy in the survey to represent all the missing galaxies at that redshift, thereby correcting the analysis and compensating for a potential bias on H_0 measurements due to the selection function. As far as I know, this is the first time that this kind of implementation is found in this way. Another methods can be seen in [6, 9] in which the approach is to "complete" the sample. Given a galaxy sample, they separate P_{cat} into two parts, the galaxy contribution and the P_{miss} contribution, tied by the relation $P_{cat} = P_{gal} + P_{miss}$. Various techniques are employed to estimate P_{miss} and compensate for the incompleteness. Both methods rely on the knowledge of a "true" distribution of the galaxies. Since this true distribution is based on the homogeneity and isotropy of the Universe, the base assumption is solid. The big trade-off between those two methods is more in the complexity of the code. having to calculate only some ratio between bins is simple and has proven to be quite successful in the final result. For test, a complete Uniform Universe is assumed, unless otherwise stated. For the final results, a non Uniform and non-complete distribution of galaxies is taken into consideration, and the incompleteness is compensated as described above.

In the above discussion on the galaxy catalogue, a detail was omitted. Theoretically, $\Pi(\theta|H_0)$ should represent both the probability of a galaxy presence and the likelihood of a merger occurring. This implies that P_{cat} should be combined with the probability of a merger at redshift z . The correct approach to computing this new prior would be to multiply P_{cat} with the previously discussed merger rate, R_{GW} . However, to avoid bias in the measurement of H_0 , this work omits the merger rate contribution and maintains consistency in this choice. Fig. 4.3 illustrates the bias using just 30 GW events and a hypothetical galaxy catalogue with the same selection function as for the Flagship. In this case, the host catalogue is extracted from a uniform distribution of host. For this test, we can forget the physical characteristics of the DSs as we are interested in checking for the presence of a bias. Figure Fig. 4.3 shows precisely the consequences of inappropriately using a catalogue with a selection function $N(z)$ without weighting the contributions as described above.

4.1.3 The Beta normalisation

We focus now our attention towards the function $\beta(H_0)$, for which a first computation was conducted in [105], assuming a Uniform universe.

β is also known as the "normalisation" because it is mathematically defined through the integral over all variables in θ , ensuring that eq. (4.12) behaves as a probability density function (PDF) for H_0 . In this context, observed luminosity distances are already taken into account. Consequently, for mock data, the observed Dl , determined from Gaussian scattering centred on the true distance (see in the next chapter) should be utilised.

The function $\beta(H_0)$ can then be defined as

$$\beta(H_0) = \int_{\mathcal{D}} P_{det}[Dl(z, H_0), \Omega, \theta'] P_{cat}(z, \Omega) P_{\theta'}(\theta') d\theta' dz d\Omega, \quad (4.18)$$

where \mathcal{D} is a region in the parameter space, $\{z, \Omega, \theta'\}$, encompassing all possible

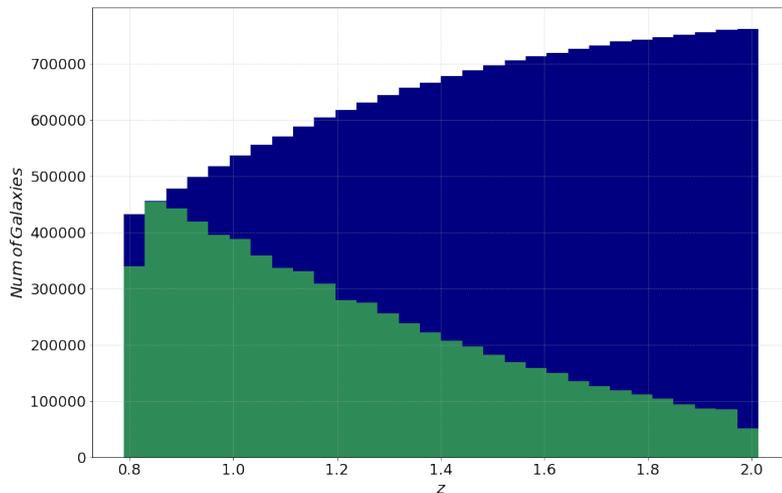


Figure 4.2. A representation of the uniform Universe, in blue, and the observed Universe, in green. The observed Universe is derived using Euclid mock data.

detectable DS.

One can observe that $\beta(H_0)$ is explicitly a function of the Hubble constant, but it also depends on the specific problem considered, as indicated by the integration domain \mathcal{D} and the galaxy catalogue prior. In real-world scenarios, β accounts for the specific interferometer in use, since each detector observes DS in different regions of the parameter space. This concept becomes clearer when considering a monochromatic source. For a given chirp mass M_c , each interferometer can detect a signal up to a minimum flux, leading to the definition of a horizon, Dl_{hor} , within which every source is detected. In this context, P_{det} is represented by the Heaviside distribution $\Theta(Dl_{hor} - Dl)$. While it is feasible to define an instrument horizon for NS due to their relatively narrow mass range, it is not possible for BH, whose mass distribution spans several orders of magnitude. A more detailed discussion on this topic is presented in [6]. It is worth noting that this remains an active area of research, with new techniques emerging regularly. A proposed strategy is to conduct a full Monte Carlo (MC) estimation of β , including source properties, particularly the masses [9]. While this method is robust, it is impractical for this study, given the vast number of potential hosts in the Euclid mock catalogue.

Instead of performing such an MC evaluation of the normalisation function, this study adopts a more physical approach, building on the work of [105, 6].

The role of $\beta(H_0)$ is to account for the implicit dependence of the likelihood on the Hubble constant. To illustrate this, consider an infinite spherical region \mathcal{R} , uniformly filled with a finite and constant density of galaxies. Assuming each galaxy has an equal probability of hosting a DS event, a box containing a certain probability of finding the DS inside it can be constructed based on the Dl of the DS and the $2D$ -sky localisation. Allowing H_0 to vary, causes the region \mathcal{R} to contract,

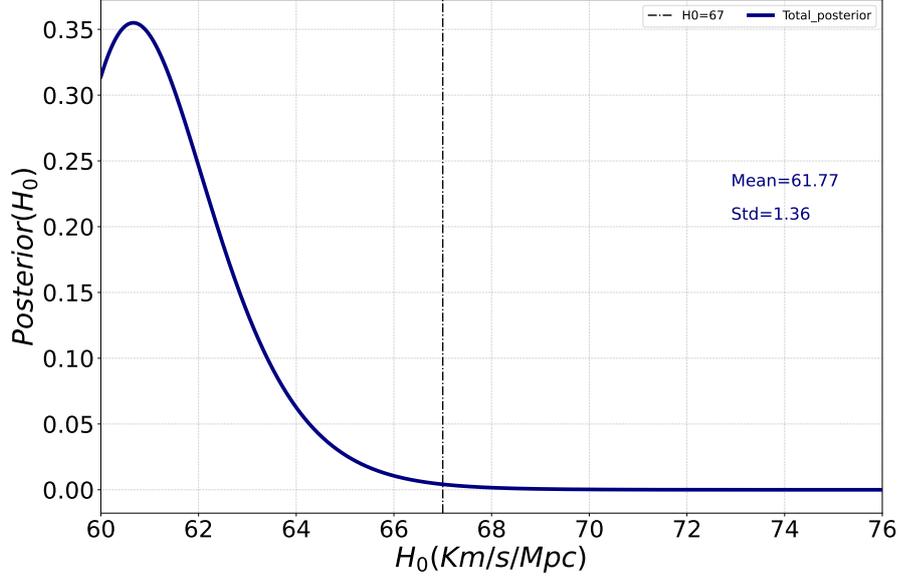


Figure 4.3. This plot demonstrates the effects of not applying the weight in a catalogue with a specific $N(z)$. The sample of DS was chosen uniformly up to a maximum distance. The exhibited bias towards lower H_0 originates from the fact that the majority of potential hosts are at "low" z , thus requiring smaller values of H_0 to satisfy $cz/H_0 = Dl$

or, from another perspective, allowing H_0 to vary causes the box to stretch and slide over the radial direction according to the inverse relation $z(Dl, H_0)$, becoming larger the farther it is from the origin. As the density in the region \mathcal{R} is fixed, there will be more potential hosts at larger distances, and since each galaxy has the same probability of hosting a DS, the likelihood will increase for higher H_0 values due to this volume effect. Essentially, this scenario deals with a homogeneous Malmquist bias, and $\beta(H_0)$ must compensate for it.

Given the above argument, it is no surprise that β is connected to the explored volume. In the following, first, the homogeneous case will be introduced, assuming an infinite galaxy catalogue, and then a method to compute β for a finite catalogue will be presented. The infinite catalogue scenario reflects the situation where the host catalogue covers a much wider region than the "box", for each value of H_0 considered.

The Homogeneous Case

I now focus on the study of the homogeneous case for β , particularly its indirect dependency on H_0 .

By definition,

$$\beta(H_0) = \int_{\Omega} \int_{z_{\mathcal{R}}} P_{det}(Dl) P_{cat}(z, \tilde{\Omega}) dz d\tilde{\Omega}, \quad (4.19)$$

where the integration over all other parameters in θ' is performed, and $z_{\mathcal{R}}$ represents the extension of the \mathcal{R} -region, which should be significantly larger than the research

volume.

P_{det} is the probability of detecting a source at a certain distance. For simplicity, I assume $P_{det}(Dl) = 1$ between a Dl_{min} and Dl_{max} and zero otherwise, implying a perfect detector within a specific Dl range. The next assumption is the homogeneity of P_{cat} in comoving volume, implemented as

$$P_{cat}(z, \Omega) = \frac{1}{V_c(z_{\mathcal{R}})} \frac{dV_c}{dzd\tilde{\Omega}}. \quad (4.20)$$

With these assumptions, eq. (4.19) can be rewritten as

$$\beta(H_0) = \frac{1}{V_c(z_{\mathcal{R}})} \int_{\Omega} \int_0^{z_{max}} P_{det} \frac{dV_c}{dzd\tilde{\Omega}} dz d\tilde{\Omega}. \quad (4.21)$$

The volume element in the innermost integral can be divided into two parts, one that depends on $\tilde{\Omega}$ alone and the other on z alone. Since one is interested in the relationship between β and redshift, the integration over the angular variables is performed.

To express the volume element, one can first define the following quantities:

$$E(z) := \sqrt{\Omega_m(1+z)^3 + \Omega_r(1+z)^4 + \Omega_k(1+z)^2 + \Omega_\Lambda}, \quad (4.22)$$

$$u(z) := \int_0^z \frac{dx}{E(x)}, \quad (4.23)$$

$$dc(z) = \frac{1}{H_0} u(z). \quad (4.24)$$

Using these definitions, the volume element is expressed as

$$\frac{d(V_c)}{dzd\tilde{\Omega}} = \frac{d_c^2 d(d_c) d\Omega}{dzd\tilde{\Omega}} = d_c^2 \frac{d(d_c)}{dz} = \frac{1}{H_0^3} \frac{u(z)^2}{E(z)}. \quad (4.25)$$

Therefore, the volume $V_c(z_{\mathcal{R}})$ is

$$V_c(z_{\mathcal{R}}) = \frac{1}{H_0^3} \int_0^{z_{\mathcal{R}}} \frac{u(z)^2}{E(z)} dz. \quad (4.26)$$

The assumption of a perfect detector is now implemented. This implies that the integral over the \mathcal{R} -region is truncated in redshift by P_{det} . Importantly, the detection probability is a function of the luminosity distance, not of the redshift, therefore β depends on H_0 through the inverse function $z(Dl, H_0)$. Moreover, given eq. (4.25), the explicit dependence on H_0 cancels out in P_{cat} .

To highlight that Dl is a variable, $\beta(H_0; Dl)$ is written as

$$\beta(H_0; Dl_{max}; Dl_{min}) = \frac{1}{\int_0^{z_{\mathcal{R}}} \frac{u(z)^2}{E(z)} dz} \int_{z_{min}}^{z_{max}} \frac{u(z)^2}{E(z)} dz, \quad (4.27)$$

where $z_{max(min)}$ is defined by $z_{max(min)} = z(Dl_{max(min)}, H_0)$.

Looking at eq. (4.26) and eq. (4.27), is possible to define the dimensionless comoving volume \mathcal{V}_c , leading to,

$$\beta(H_0, z_{max}, z_{min}) = \frac{\mathcal{V}_c(H_0, z_{max}, z_{min})}{\mathcal{V}_c(H_0, z_{\mathcal{R}})} \quad (4.28)$$

with the denominator representing the volume of the catalogue.

Eq. (4.28) is derived under the assumption of a homogeneous catalogue. It will be shown later that, when one turns on the statistical weights $w(z)$, this formula is equally applicable to an observed galaxy catalogue with an arbitrary selection function. The equation for β is also well suited for fast computation. In a homogeneous Universe, the number of galaxies is proportional to the volume, simplifying eq. (4.28) to a sum or weighted sum, when the $w(z)$ are present, over the catalogue:

$$\beta(H_0, z_{max}, z_{min}) = \frac{1}{\sum_{z_\alpha \in \mathcal{R}} w(z_\alpha)} \sum_{z_{min} \leq z_\alpha \leq z_{max}} w(z_\alpha). \quad (4.29)$$

This equation effectively compares the volume of the box containing a certain probability of finding the DS inside it with the volume of the catalogue when $w(z) = 1$.

It is already stated that eq. (4.28) and eq. (4.29) are dependent on H_0 only via the inverse relationship $z = z(Dl, H_0)$. The focus is now to understand the general behaviour of this normalisation as a function of H_0 and the integration limits.

The denominator of eq. (4.27) is constant, representing the volume of the catalogue, denoted as K . This K is the sum of three terms: the integration from 0 to z_{min} , from z_{min} to z_{max} , and from z_{max} to $z_{\mathcal{R}}$, defined as A , B , and C respectively. Hence, $\beta = B/(A + B + C)$, illustrating that β is always less than or equal to one.

The function

$$j(z) = u(z)^2/E(z) \quad (4.30)$$

is now defined. Since $u(x) > 0$ for all $x \in \mathbb{R}^+$ and the same holds for E , $E(x) > 0$ for all $x \in \mathbb{R}^+$, implies that the integrand, $j(z)$, is a positive function in z , making β a non-decreasing function of z via the integration extremes. This implies specific limits for the range in which β can vary.

For z_{min} approaching 0, $A \rightarrow 0$, so, since β is a continuous function, it must be true that $\beta = B/(B + C) := m$. As z_{max} approaches $z_{\mathcal{R}}$, $C \rightarrow 0$, leading to $\beta = B/(A + B) := M$, where $0 < m < M \leq 1$. This is evident when setting $z_{min} = 0$ from the start, finding $M = 1$ while m is a constant approaching 0 as z_{min} approaches 0.

The study of β as a function of z is insightful for understanding the role of H_0 . It will be shown that a function, increasing continuously between 0 and 1 with an initially positive second derivative, must have a single inflection point.

Given $\mathcal{V}_c = \int j(z)dz$, and with $\mathcal{K} = K/H_0^3$, the normalisation is can be cast as

$$\beta = \frac{\mathcal{V}_c(z_{min}, z_{max})}{\mathcal{K}}. \quad (4.31)$$

Focusing on the simple case $z_{min} = 0$, eq. (4.31) is investigated as a function of the Hubble constant. The relationship $z(Dl, H_0)$ is now denoted as $\zeta(dl, h)$ to have a lighter notation, where $dl \equiv Dl$ and $h \equiv H_0$.

To understand how β behaves as a function of h , its first and second derivatives in h are studied using the *Leibniz integral rule*, which, assuming that all the involved functions are smooth enough to perform such operations, is

$$\frac{d}{dx} \int_{a(x)}^{b(x)} f(x, t) dt = \int_{a(x)}^{b(x)} \partial_x f(x, t) dt + f(x, b(x))b'(x) - f(x, a(x))a'(x). \quad (4.32)$$

Even if the denominator of β is constant in z , it will not be constant once we allow H_0 to vary. The focus now shifts to the numerator, as the calculations are similar. It is important to note that when performing the derivative with respect to h , the derivation rule for the ratio of two functions must be applied. It is also worth mentioning that this computation is highly complex, and a detailed description of each step would result in an excessively long text that might not offer more insight than the final plot in Fig. 4.4. Therefore, only key parts of the demonstration will be provided here.

For simplicity, by setting $z_{min} = 0$, eq. (4.31) can be reformulated as

$$\frac{\mathcal{V}_c}{\mathcal{K}} = \frac{\int_0^{\zeta(dl^*, h)} j(x) dx}{\int_0^{\zeta(\bar{dl}, h)} j(x) dx}, \quad (4.33)$$

As previously stated, \mathcal{V}_c does not explicitly depend on h , hence, the significant parts in the calculation, according to the Leibniz rule, are those proportional to

$$\frac{d^n \zeta(dl^*, h)}{d h^n}. \quad (4.34)$$

It is emphasised that once $dl = dl^*$ is fixed, $\zeta(dl^*, h)$ essentially becomes a function of h alone.

ζ is now explicitly defined using the *deceleration* parameter q . Considering only terms up to second order, the inverse relation $z - dl$ is given by

$$z = \zeta(dl, h) = dl h + \frac{1-q}{2} dl^2 h^2 + O(dl^3 h^3). \quad (4.35)$$

The deceleration parameter q , linked to the Hubble parameter, evolves with the Universe and is described by

$$q(z) = -\frac{\ddot{a}(z)a(z)}{\dot{a}^2(z)} = \frac{1+z}{E(z)} \frac{dE(z)}{dz} - 1, \quad (4.36)$$

thus, for consistency with the second-order approximation, q is taken to be fixed

equal to q_0 , i.e. its value at present time.

The complete equation for the second derivative of β with respect to h reads

$$\frac{d^2}{dh^2}\beta(h) = \frac{1}{\mathcal{K}} \frac{d^2}{dh^2}\mathcal{V}_c - \frac{\mathcal{V}_c}{\mathcal{K}^2} \frac{d^2}{dh^2}\mathcal{K} - \frac{2}{\mathcal{K}^3} \frac{d}{dh}\mathcal{K} \left(\mathcal{K} \frac{d}{dh}\mathcal{V}_c - \mathcal{V}_c \frac{d}{dh}\mathcal{K} \right). \quad (4.37)$$

Recognising that the volume at the denominator of eq. (4.37) is substantially larger than in the numerator, the most significant term is therefore the one being proportional to $\frac{1}{\mathcal{K}}$:

$$\frac{1}{\mathcal{K}} \frac{d^2}{dh^2}\mathcal{V}_c. \quad (4.38)$$

The second derivative of \mathcal{V}_c is found to be proportional to $dl^{*2} - dl^{*2}q$. Given that the deceleration parameter at $z = 0$ is known to be negative, it follows that the second derivative is positive, indicating that β is a convex function and grows as a cubic function, as established in [105, 107] for the low redshift case.

Upon increasing dl^* , and consequently \mathcal{V}_c , the assumption $\mathcal{V}_c \ll \mathcal{K}$ weakens. Higher-order terms, proportional to $1/\mathcal{K}^2$, must therefore be considered. These new contributions are negative and grow faster than the first term, since the volume increases more rapidly than its derivative. This leads to the conclusion that there must exist a value of h where the second derivative changes sign and turns negative, making β concave. The term proportional to $1/\mathcal{K}^3$, while positive, does not increase as quickly, so the equation (4.37) remains negative once reached the inflection point. The critical insight is that $d^2\beta/dh^2$ has only one root, meaning the function remains concave but still always increasing as $z = \zeta(dl, h)$ increases, reaching an asymptotic value of 1 only when $z_{max} \rightarrow \infty$, and consequently $dl^* \rightarrow \infty$. If z_{min} is not zero, the same procedure can be applied, by simply inverting the extremes of integration and keeping z_{max} fixed. This procedure leads to the conclusion that the second derivative is always positive in this case, and β remains convex.

This result, which holds whenever a maximum distance dl^* can be defined, is not only counterintuitive but also very general. Its certainty applies to NSNS type events, since their mass range is narrow and therefore a dl^* can be well-defined. For BHBH events, an effective dl^* can be introduced in two main ways: either by selecting an efficiency threshold for the interferometer, i.e. by setting a certain percentage threshold of detection of the total possible events (noting this requires knowledge of the intrinsic distribution of sources), or by dividing BHBH events into chirp-mass bins. Thus, for each bin i , a dl_i^* can be defined following the aforementioned results. The total posterior can then be computed by combining every i -th posterior.

The Homogeneous Case: Finite Size

Up to now, I have assumed that, for every tested value of H_0 , the galaxy catalogue is substantially larger than the box containing a certain probability of finding the DS inside it. This holds true when using data from early GWs observations, but may not be the case with next-generation interferometers.

To account for the boundaries of the catalogue, an *effective* box volume is first

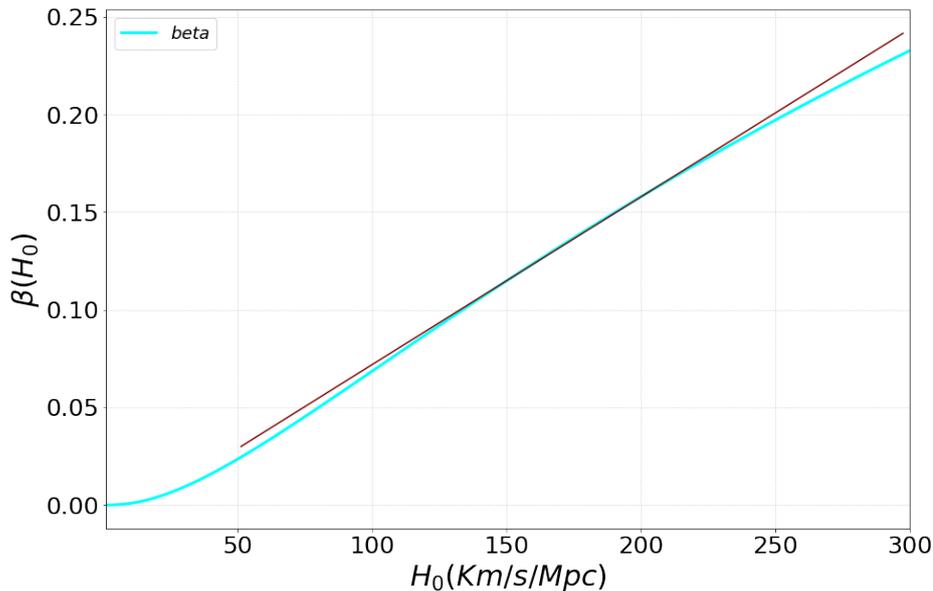


Figure 4.4. The full evolution of the function $\beta(H_0)$. The selected dl^* is 10.4 Gpc. The red line helps visualise the change of regime from convex to concave.

defined. The maximum and minimum z of the catalogue are denoted as \bar{z} and \underline{z} , respectively. The concept is that if the box extends outside the catalogue, for instance, $z_{max} > \bar{z}$, then z_{max} is replaced by \bar{z} , with a similar adjustment for z_{min} . Therefore, we have

$$\beta(H_0, z_{min}, z_{max}) = \frac{\mathcal{V}_c(z_{min}, z_{max})}{\mathcal{V}_c(\underline{z}, \bar{z})}, \quad (4.39)$$

along with the definitions for z_{min} and z_{max}

$$z_{min} = \begin{cases} z(H_0, Dl_{min}); & \text{if } z(H_0, Dl_{min}) \geq \underline{z} \\ \underline{z}; & \text{if } z(H_0, Dl_{min}) < \underline{z} \end{cases} \quad (4.40)$$

$$z_{max} = \begin{cases} z(H_0, Dl_{max}); & \text{if } z(H_0, Dl_{max}) \leq \bar{z} \\ \bar{z}; & \text{if } z(H_0, Dl_{max}) > \bar{z} \end{cases} \quad (4.41)$$

This approach effectively addresses the finite size issue, particularly when a DS, for each tested value of H_0 , does not go out from the catalogue boundary, but the box does. However, a comprehensive solution to this problem is still needed. Another potential solution is proposed in [103]. It is important to note that cross-correlating DS that do not share a footprint with a host catalogue is far from ideal. The limitation of this technique is that the DS must lie within the redshift range of the catalogue. While it is acceptable for a neighbourhood of the DS to exceed the boundaries, the correction fails if the DS is entirely outside the host catalogue.

4.1.4 Analytical Insights

In this subsection, the aim is to present some analytical results that provide insight into what can be expected from the posterior over H_0 in certain simple cases.

The analysis relies on the following assumptions:

- The catalogue is complete.
- The uncertainty on the luminosity distance is isotropic.
- The normalisation term $\mathcal{N}_i(\Omega)$ is isotropic.
- The precision of the host localisation is infinitely larger than for a DS.
- Without loss of generality, $z = 0$ is set.
- The hosts catalogue is significantly larger than every possible research box.

Under these assumptions, the function $\beta(H_0)$ can be expressed as

$$\beta(H_0, Dl_{max}) = \frac{\sum_{\alpha; z_{\alpha} < z_{max}} w_{\alpha}}{\sum_{\alpha}^{N_{cat}(\mathcal{R})} w_{\alpha}}, \quad (4.42)$$

where $w_{\alpha} = w(z_{\alpha})$ and $N_{cat}(\mathcal{R})$ is the number of galaxies in the region. From eq. (4.11) and eq. (4.12):

$$\begin{aligned} \mathcal{L}(\{D_{GW}\} | H_0) &= \\ &= \prod_{i=1}^{N_{obs}} \frac{\sum_{z_{\alpha} < z_{max}}^{N_{cat}(\mathcal{R})} w_{\alpha}}{\sum_{z_{\alpha} < z_{max}} w_{\alpha}} \int_{\mathcal{R}} dz d\Omega \frac{\rho_i(\Omega) \mathcal{N}_i}{\sqrt{2\pi}\sigma_i} e^{-\frac{(Dl(z, H_0) - \mu_i)^2}{2\sigma_i^2}} \frac{\sum_{\alpha}^{N_{cat}(\mathcal{R})} w_{\alpha} \delta(z - z_{\alpha}) \delta(\Omega - \Omega_{\alpha})}{\sum_{\alpha}^{N_{cat}(\mathcal{R})} w_{\alpha}}. \end{aligned} \quad (4.43)$$

Since the statistical weights w_{α} are fixed in z , they can be moved out of the integral, simplifying the equation to

$$\mathcal{L}(\{D_{GW}\} | H_0) = \prod_{i=1}^{N_{obs}} \frac{1}{\sum_{\alpha; z < z_{max}} w_{\alpha}} \sum_{\alpha}^{N_{cat}(\mathcal{R})} \frac{\rho_i(\Omega_{\alpha}) \mathcal{N}_i}{\sqrt{2\pi}\sigma_i} e^{-\frac{(Dl(z_{\alpha}, H_0) - \mu_i)^2}{2\sigma_i^2}} w_{\alpha}. \quad (4.44)$$

For simplicity, all constants are grouped together under the symbol Γ , and only a single GW event is analysed, hence $N_{obs} = 1$.

$$\mathcal{L}(\{D_{GW}\} | H_0) = \frac{\Gamma}{\sum_{z_{\alpha} < z_{max}} w_{\alpha}} \sum_{\alpha}^{N_{cat}(\mathcal{R})} \rho(\Omega_{\alpha}) w_{\alpha} e^{-\frac{(Dl(z_{\alpha}, H_0) - \mu)^2}{2\sigma^2}}, \quad (4.45)$$

or more explicitly:

$$\begin{aligned} \mathcal{L}(\{D_{GW}\} | H_0) &= \\ &= \frac{\Gamma}{\sum_{z_{\alpha} < z_{max}} w_{\alpha}} \sum_{\alpha}^{N_{cat}(\mathcal{R})} \rho(\Omega_{\alpha}) w_{\alpha} \exp \left\{ - \left(\frac{1 + z_{\alpha}}{H_0} \int_0^{z_{\alpha}} \frac{dx}{E(x)} - \mu \right)^2 / 2\sigma^2 \right\}. \end{aligned} \quad (4.46)$$

Eq. (4.46) shows that the posterior for a single event comprises the sum of a number N_{cat} of quasi-Gaussian shaped functions. While each function is not strictly a Gaussian, since eq. (4.46) is not a Gaussian function with respect to the variable H_0 , in the infinite precision case they essentially behave like Gaussian functions. Fig. 4.5 illustrates the result for a simplistic scenario: one host and one DS with a 1% error over the luminosity distance.

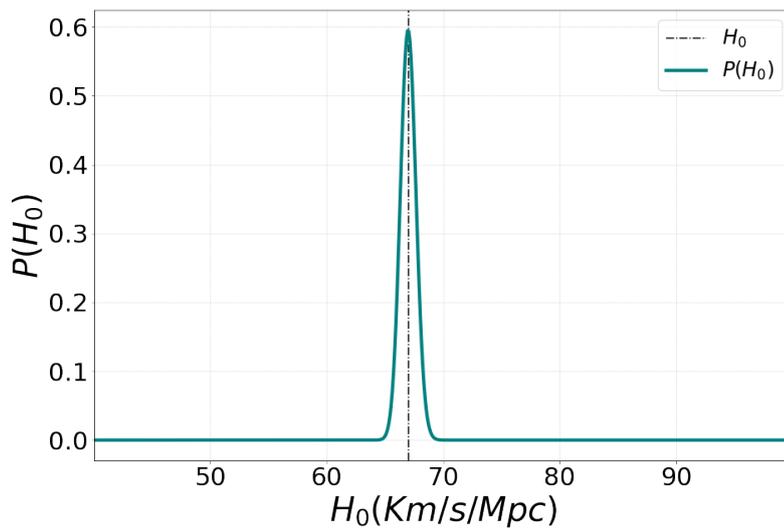


Figure 4.5. The posterior for a single DS and a single host. The curve is centred at the correct value of H_0 . This figure shows that for small errors over μ , the observed luminosity distance, the posterior can be well approximated by a Gaussian function.

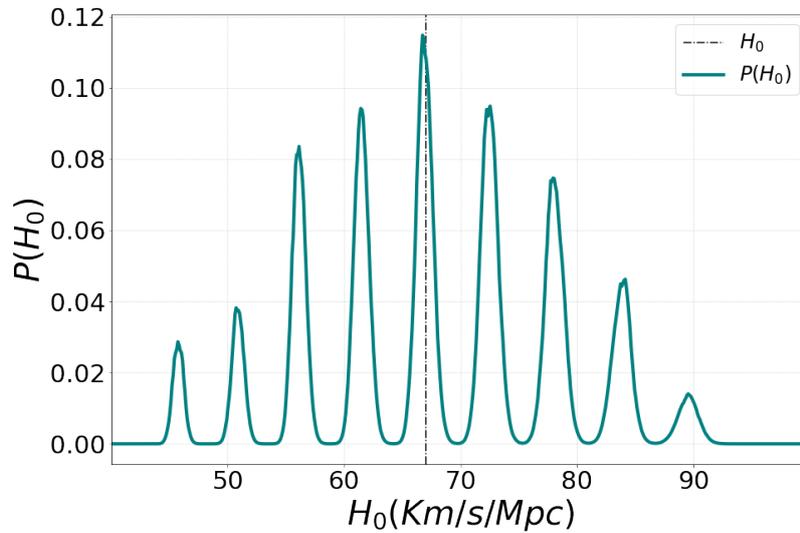


Figure 4.6. The posterior for a set of host and a single DS. The hosts are placed in such a way that their contribution to the overall posterior do not overlap. Each host is placed on a different line of sight and at a different redshift. The plot shows that in this case, the highest peak belongs to the “true” host. We can observe the effect of ρ on the other peaks.

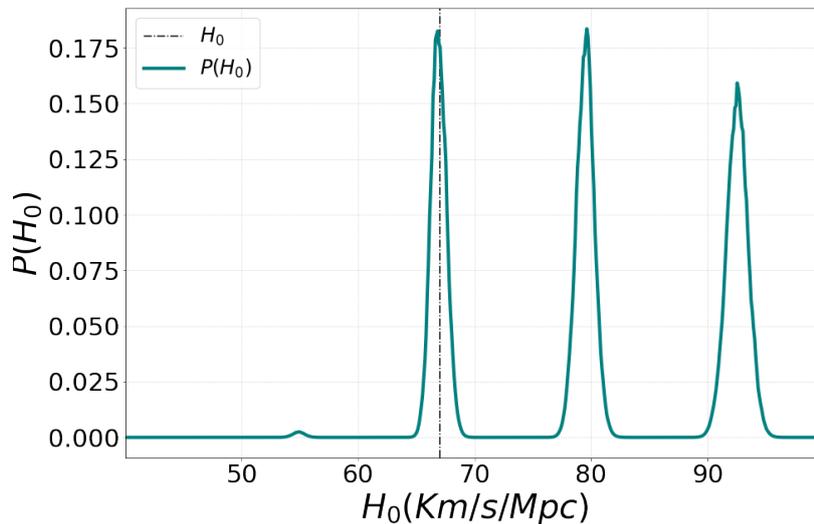


Figure 4.7. The posterior for a single DS and a set of hosts. The hosts are placed along the same line of sight, except for the one contributing to the small bump at $H_0 \sim 55$. This plot shows both the widening of each individual peak due to the $1/H_0$ factor and the fact that all peaks along the same line of sight would roughly have the same height if not for the normalisation.

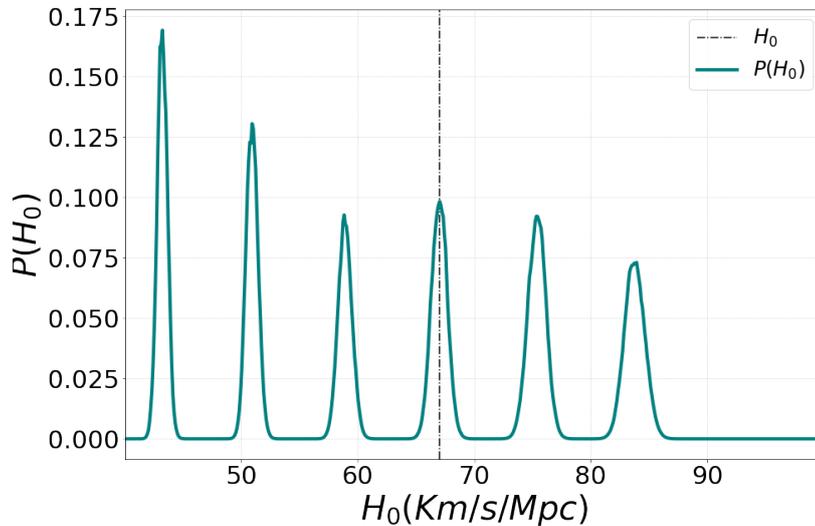


Figure 4.8. The posterior for a single DS and a set of hosts both at different redshifts and along different lines of sights. Some parts of the catalogues are more populated, placing more hosts in a narrow region.

Considering the angular probability ρ , when there are few possible hosts, the one in the correct location, that is, the host with the same RA and DEC of the DS, leads to a higher peak in the final posterior. When fixing the angular probability ρ , it can be noted that, if H_0 spans a wide range of values and if the collection of z_α are roughly of the same order, the value of the integral in eq. (4.46) remains relatively constant. This is because, independently of the redshift, a specific H_0 can always be found such that it satisfies the condition $\mathcal{D}l/H_0 - \mu = 0$, where $\mathcal{D}l$ represents $(1 + z_\alpha) \int_0^{z_\alpha} dx/E(x)$. Essentially, this is the case where $\mathcal{D}l/H_0^{max} - \mu \ll 0$ and $\mathcal{D}l/H_0^{min} - \mu \gg 0$ for all z_α .

Without normalisation, all the peaks would have the same height. Additionally, given the form of eq. (4.46), it is expected that peaks at large values of H_0 will be broader due to the $1/H_0$ behaviour of the exponential. It is only thanks to the normalisation, acting upon those wider peak, that the amplitude is correctly reduced. Both effects can be observed in Fig. 4.6 and Fig. 4.7, where in the former the hosts are placed at various redshifts and along multiple lines of sight, resulting in a bell-shaped total posterior that follows the profile of the distribution $\rho(\Omega)$, while the latter plot demonstrates how two “close” peaks can have the same height and also exhibits the increasing spread of a single peak over H_0 due to the $1/H_0$ in the exponent.

The role of the redshift is more subtle, if the catalogue is denser for some redshift values, the peaks from those hosts will sum up together, building up the likelihood for that value of H_0 and this can lead to peaks that are higher than the “true” one. This is shown in Fig. 4.8 in which the starting point is a configuration similar to the one in Fig. 4.6 but with the addition of more possible hosts for a specific value

of redshift. While each individual host contributes with a weaker peak, since they overlap, they result in a much higher contribution for $H_0 \sim 52$ and $H_0 \sim 40$.

All the plots above help us to finally explain the aforementioned white noise behaviour for the peaks other than the one for the “true” value of H_0 . As shown, every time $\mathcal{D}l/H_0 - \mu = 0$ is satisfied, each event returns a combination of peaks. A single event, in a realistic case, will produce a peak for each galaxy in the box, with the height modulated by the angular distribution and the normalisation. Since there is only one true value of H_0 , When the individual posteriors are multiplied, the only contribution that consistently adds up is the one corresponding to the true value of H_0 . This is because it is the only value that can generate a peak in each posterior simultaneously. The convergence speed of this process depends on various factors, such as the density of the host catalogue, the uncertainty over the angular probability and the uncertainty over the luminosity distance of the GW. The essence of this method lies in the fact that, given a certain number of DSs, the uncertainty over H_0 will become competitive with other cosmological probes.

All of the above results are based on certain assumptions, which are best disclosed in this final paragraph. As can be seen in eq. (4.13) we are assuming that the posterior on Dl is Gaussian; this is a simplification, but it is a good starting point. The function ρ , as mentioned, is a multivariate Gaussian centred in the angular coordinates of DS. CODE does not take into account the correlation between the error on distance and the error on angular position, a correlation that is present in reality. CODE, apart from the statistical weights discussed earlier, does not assign probabilities based on the brightness or mass of the host galaxy, so every galaxy in the search region have the same probability to be the host of the source.

Chapter 5

The CODE Pipeline

The Cross correlation with Dark sirEns (CODE) pipeline is designed to generate, combine and analyse any two different types of catalogues: a spectroscopic collection of galaxies and an ensemble of DSs.

As described so far, CODE has been designed to perform a Bayesian, cross-correlation analysis between an Euclid-like spectroscopic catalogue of galaxies that may host, or correlated with, a catalogue of DS detected by future observational campaigns performed with the ET.

In the current implementation, and in absence of real datasets, the input DS and galaxy catalogues consists of mock objects. These catalogues can be either imported or, which is the case considered in this Thesis, internally generated. Catalogues of real objects can, in principle, be also used once they will be available. When generating the simulated samples, I have made some simplifying assumption on the shape of the DS detection area. The result footprints do not quite match those observed so far in the O3 and O4 runs. But are simple and realistic enough to simulate expectations for future datasets like those obtained from ET.

The core of the CODE pipeline performs the cross-correlation of the objects in the two catalogues in order to estimate cosmological parameters through a hierarchical Bayesian analysis. If the host catalogue is not complete, which is always the case with realistic samples, then the statistical weights derived from the observed redshift distribution $N(z)$ need to be included to restore statistical uniformity.

The outcome of the analysis is the posterior distribution of the cosmological parameters, in the current implementation I only consider H_0 . I also provide the normalising function $\beta(H_0)$ and the sub-catalogue of DS used to perform each specific analysis, so that the same DSS could be reprocessed in a different analysis for comparison purposes, if required. Finally, CODE produces also the likelihood for each DSs analysed, still not divided by β

CODE has the same scientific goals and shares several similarities with other existing codes, such as those developed by [6, 9]. However, I have put more emphasis on the modularity of the code and made an effort to make it as user-friendly as possible.

The CODE package also includes a uniform host generator, designed for calibration and testing, and to provide users with an ideal starting catalogue to play

with. CODE then takes the host catalogue, either the realistic or the ideal one, to generate the parted DS catalogue which specifies the physical properties of the DS from which the “observed” GW sample is obtained. Fig. 5.1 depicts the main steps of the analysis. Sample selection and creation of the ‘observed’ catalogues are represented by the two rotated squares. Once the catalogues are produced, Bayesian hierarchical inference is operated to evaluate the posterior probability function of the sought-out parameters, in this case, H_0 .

This chapter will describe in detail the CODE pipeline with emphasis on the generation of the “intrinsic” i.e., instrument-independent catalogues.

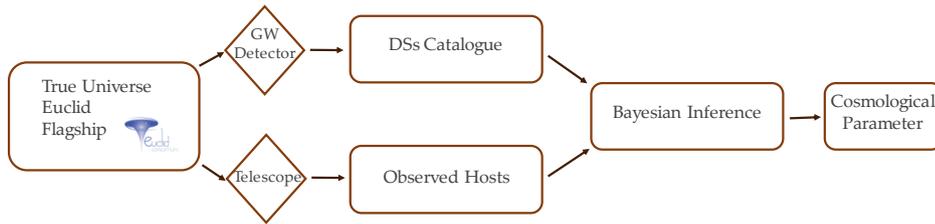


Figure 5.1. The image shows the coarse-grained structure of the CODE pipeline. The starting point is the parent catalogue, then an instrument is chosen to observe both for the DSs and the hosts. Having both catalogues, it is possible to use Bayesian inference to estimate cosmological parameters.

5.1 The Parent Catalogues

The starting point of the pipeline is the parent catalogue. This first step, manipulation of the parent catalogue, is really relevant when the pipeline is used in the forecast mode, i.e. when simulated data are used to make predictions and/or to optimise the analysis. The analysis of real data, when this will be available, really starts from the second step, since in that case both the DS and the galaxy catalogues will be available.

Since there is no real data available yet, a parent sample, dubbed “true Universe” in the flow chart, is first needed. From the parent sample, both the host galaxies and the DSs samples will be “observed”. In this chapter, I will use two parent samples: the Flagship simulation of the Euclid spectroscopic sample and a sample of objects uniformly distributed within a pre-defined volume. The latter catalogue, is of course ideal, and is used to perform validation tests and sanity checks.

A fundamental requirement for the true Universe sample is that it must contain significantly more objects than those in the catalogues that will be extracted from, to properly include the various selection effects one needs to account for. A second, related issue, is the relative number of DSs and galaxy hosts. The Bayesian analysis implemented here assumes that all DSs events are uncorrelated. This is strictly true when each DSs is drawn from an independent realisation of the galaxy catalogue. It is customary to assume that this hypothesis is satisfied also when a single galaxy

catalogue is available, like in the case of real data, and the number of DS is much lower than that of galaxies so that correlation between the DSs events can be ignored. I shall also make this assumption since, in all cases considered, host galaxies outnumber the DSs.

5.2 The Dark Sirens Catalogue

Once the parent catalogue is created, it is possible to generate a DS sample of potentially observable sources, where all the relevant intrinsic properties of the DS are specified. I will refer to this sample as the 'intrinsic DS catalogue'. The type of intrinsic DS catalogue that is generated depends on the type of parent sample used, of which we have two.

The first parent catalogue simply consists of a sample of objects homogeneously distributed within a specified volume. The geometry and the number density of the objects fully characterize this catalogue that I will refer to as the 'True Universe'. Its corresponding 'intrinsic DS catalogue' will also consist of a set of uniformly distributed sources whose number density can be modulated using a Monte Carlo rejection procedure to match any target $N_{DS}(z)$, if necessary. I do not specify the intensity of these GW sources since these are assumed to be 'golden events' that will be detected, so that no signal-to-noise ratio needs to be computed.

The second parent catalogue used is the Flagship simulation. In this case the intrinsic DS catalogue is generated by randomly selecting a sample of objects (the DS hosts) with the redshift distribution given by eq. (4.6) although, as we will see, the shape of the DSs redshift distributions does not have a strong effect on the posterior probability. Then, one DS progenitor is associated to each of these hosts. The progenitor is assumed to be a Black Hole binary system characterized by the masses of the two BHs, inclination of the orbital plane with respect to the line of sight, two angular coordinates ϕ and θ (RA and DEC), the polarisation angle ψ and the mass ratio q . The Black hole masses M_1 and M_2 are drawn from the probability distribution proposed by [99]. The inclination angle ι and the polarisation are all drawn from a uniform distribution, since no preferential orientation is expected in any of these cases.

Along with these two types of intrinsic DS catalogue, a companion catalogue of potential host galaxy is also generated. In the True Universe case, the host galaxy sample is also generated by randomly resampling the parent sample, modulated to match a target $N_h(z)$, if required. In the Flagship case, since the parent catalogue is designed to mimic the Euclid spectroscopic catalogue, the parent catalogue of galaxy host coincides with the parent sample.

Making the DS catalogue. Step 1: Spatial Distribution.

The first step to create the DS catalogue is to determine the spatial coordinates of the DS. The same procedure is used for both the True Universe and the Flagship parent catalogue. First, the redshift of the DS is assigned according to eq. (4.6) using the same rate as [99]. The absolute normalization is left free. This choice

guarantees the correct relative number density of DS as a function of redshift and allows me to create a "master" catalogue with a very large number of DS from which I can draw different independent DS samples. The master catalogue is split into different redshift shells with arbitrary width.

Once the number N_i of DS in the i -th z -bin is determined, then there are two options to assign angular coordinates. The first option simply sets the angular coordinates of the DS equal to those of one host galaxy randomly selected from the companion galaxy sample in the same z -bin. With this option, the DS is guaranteed to have its host in the companion galaxy catalogue. The second option adds an additional step. Once the galaxy host is randomly selected, the DS is placed in its neighbourhood isotropically and at a radial distance drawn from a probability function designed to match the small-scale galaxy-galaxy 2-point correlation function, modelled as:

$$\mathcal{D}(r) = \left(\frac{r}{r_0}\right)^{-1.8}; \quad 0.1 \leq r \leq 10; \quad r_0 = 5 \text{ Mpc}. \quad (5.1)$$

In Fig. 5.2, the histogram shows the resulting cumulative radial distribution of DS around host galaxies in one of our DS catalogues. The red diagonal line, drawn for reference, shows the expected $r^{-0.2}$ scaling. This second step relaxes the assumption of the DS host being in the galaxy catalogues but preserves the spatial correlation between DS and galaxies, which is expected since they both trace the underlying mass distribution."

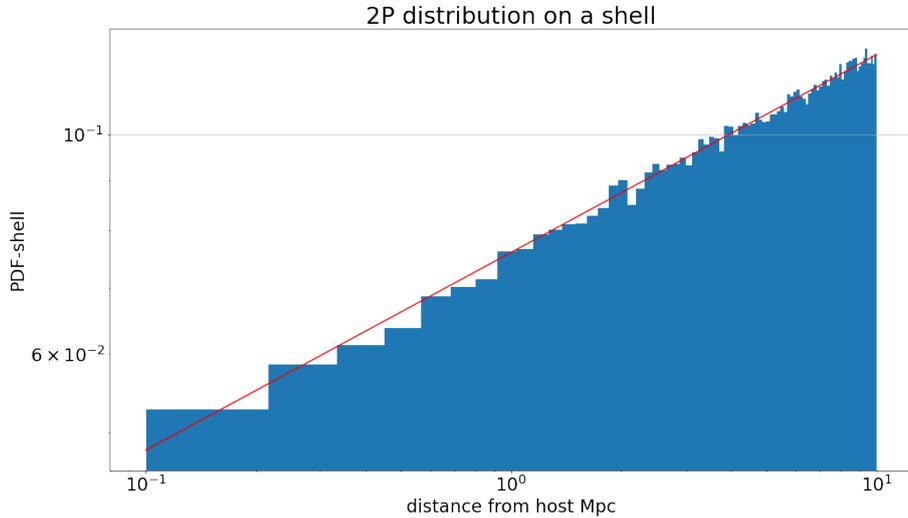


Figure 5.2. The distance distribution DS Host vs Galaxy in the catalogue. The red line is the integral of $\mathcal{D}(r)$ over a spherical shell, and the histogram is the real extraction from the catalogue. Both curves are normalised and integrate to 1.

Making the DS catalogue. Step 2: DS masses

Having established the coordinates of the DSs, the next step is to assign to the DS the intrinsic physical properties, such as masses, distances, and orientations, on which the GW signal depends.

As anticipated, the inclination of the orbital plane of the DS and the polarization angle of the GW signal are both selected from random distributions. As for the masses of the binary members, we adopt the same probability distribution functions as [99]. The primary mass M_1 is then drawn from the following composite function,

$$\mathcal{W}(M_1) \propto \begin{cases} M_1^{-1.5} & 5 \leq M_1 \leq 40 \\ M_1^{-5} & 40 \leq M_1 \leq 80 \\ 0 & 80 \leq M_1 \leq 120 \\ 1/M_1 & 120 \leq M_1 \leq 10^4, \end{cases} \quad (5.2)$$

where the mass M_1 is expressed in units of M_\odot . This phenomenological probability function is a multi-broken power law with no discontinuity except at the mass gap $80 \leq M_1 \leq 120$. Fig. 5.3 compares the mass function in one of the DS catalogues used in this Thesis (blue histogram) with the expected shape from eq. (5.2). The function, represented in logarithmic unit, spans several order of magnitudes. Despite the discreteness effects due to the paucity of objects which is evident in the high-mass tail the distribution matches theoretical expectations, as shown in the high mass tail zoom in of Fig. 5.4

The probability distribution of the q parameter is

$$P(q) \propto \begin{cases} \frac{1}{q} & M_1 \leq 80 \\ \sqrt{q} & 120 \leq M_1 \leq 10^4 \end{cases} \quad (5.3)$$

with q between 0.1 and 1 and the masses expressed in M_\odot . As described in [99] a cut-off is imposed on the secondary masses. Since the DSs are BHs pairs, the minimum secondary mass is set to $5M_\odot$ ¹. This cut-off modify the shape of the probability distribution, making the probability flat for low values of the mass ratio. Since low mass BHs are more numerous, the mass ratio distribution is dominated by the $1/q$ behaviour. In Fig. 5.5 the distribution for q is shown both in linear and double-log scale.

Making the DS catalogue. Step 3: From the Intrinsic to the Observed DS Catalogue.

The final step in the procedure involves 'observing' the intrinsic DSs sample using a specified GWs detector. In the current CODE pipeline implementation, an ET-like network comprising uncorrelated interferometers is considered. I have also included the aLIGO instrument for testing purpose only.

¹If $qM_1 < 5$, then a new q is extracted until the secondary masses are equal or greater than the cut-off

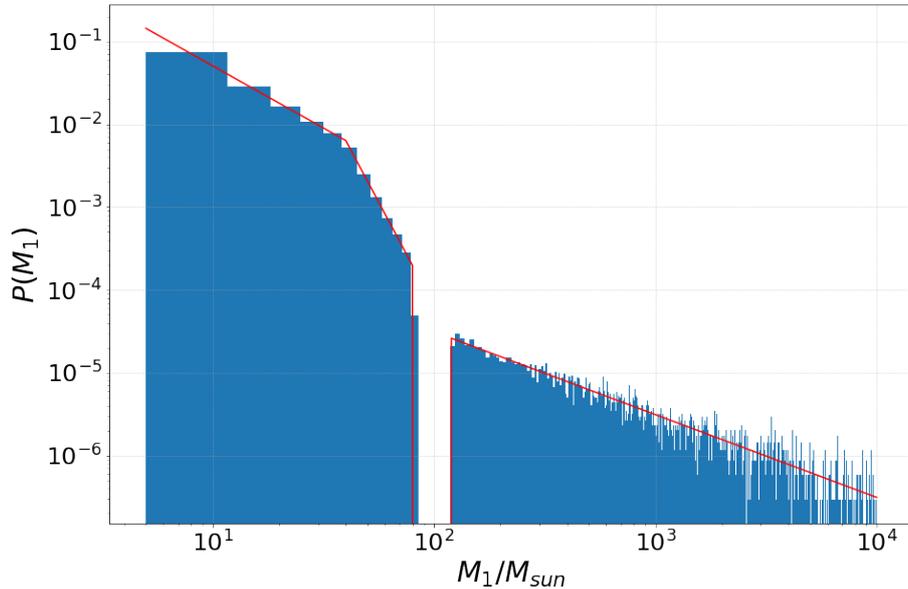


Figure 5.3. The primary mass probability distribution for the DS. The red line is $\mathcal{W}(m_1)$ and the histogram is the real extraction from the catalogue. Both curves have the area normalised to 1.

The antenna pattern is described by equation (3.18), and the total signal-to-noise ratio is given by the squared sum of the individual antenna SNRs (see equation 3.19). I have also considered, for testing purpose only, the case of the LIGO

To validate this step, we have run a specific test aimed at matching the results of [100], i.e. at reproducing the same efficiency curves for the aLIGO and ET-D interferometers. These curves have been obtained assuming a N BBH systems with a specified total mass, sky location, redshift, inclination angle and polarisation angle (both randomly oriented) and using only the + part of the GW signal. The efficiency is estimated in various redshift bins as the ratio between the number of detected DSs over the total one. The results obtained with my code matches well those in the literature, showing that the routine for evaluating the SNR is unbiased. This is especially true for aLIGO.

² The end product of the pipeline is a mock catalogue of DS, that mimics the results of an observational campaign conducted with the ET. I conclude this section by reaffirming the utilisation of both the extrinsic and intrinsic catalogues. Given the intrinsic catalogue, that is, the compendium of sources that Nature offers to us. A particular instrument, specifically ET, is now selected. Subsequently, the SNR is computed, resulting in the distribution of the observed DS. This juncture is the sole instance within the pipeline where orbital parameters are utilised. With

²This is significant as it is a well known instrument, while there are no real data for the ET case.

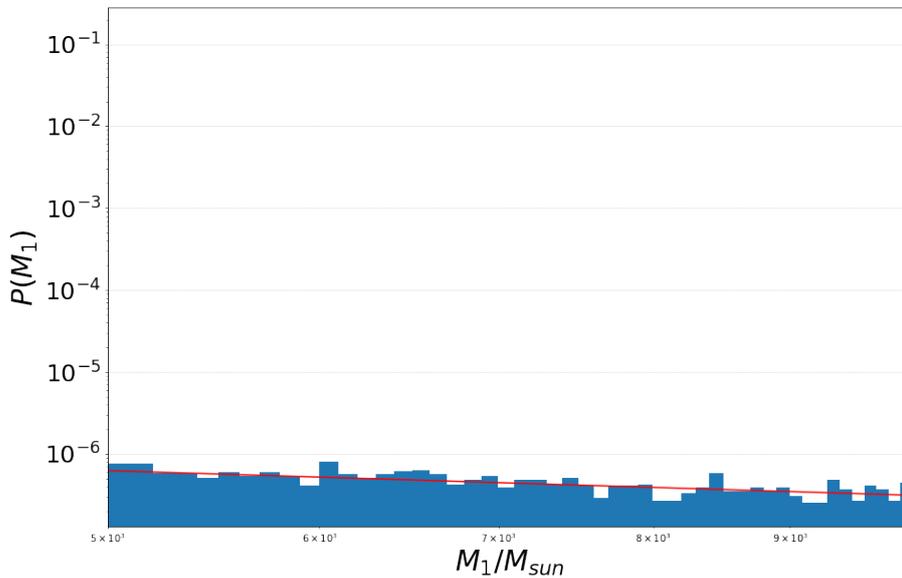


Figure 5.4. The primary mass probability distribution for the DS. The red line is $\mathcal{W}(m)$ and the histogram is the real extraction from the catalogue. This figure focuses on the high mass region of the distribution.

the SNR computed for all possible DS, it becomes possible to select subsamples possessing some desired characteristics; in this case, golden-events. In essence, the final outcomes of this thesis are derived from a distribution of 'observed' DS, and observing them necessitates the assignment of all necessary parameters to compute the signal-to-noise ratio.

5.3 Making the Host Galaxy Catalogues

Three main types of parent datasets were used to select the catalogue of galaxy hosts. The first one, which I have introduced previously, is the so-called True-Universe. It consists of a set of objects randomly distributed with a constant mean density within a given comoving volume.

The second type of host galaxy catalogue is obtained by selecting a subset of objects from the True-Universe catalogue with no angular modulation but a redshift distribution $N(z)$ matching that of the Flagship sample. Finally, the third type of host galaxy catalogue is the Flagship sample itself. Unlike the first two samples, whose objects have no intrinsic clustering, host galaxies in the Flagship have their characteristic spatial correlation properties. The first two catalogues were utilised for testing purposes. Characterised by a uniform distribution of hosts, they represent the worst-case scenario for inferring H_0 , as the absence of cross-correlation between DSs and hosts broadens the posterior distribution.

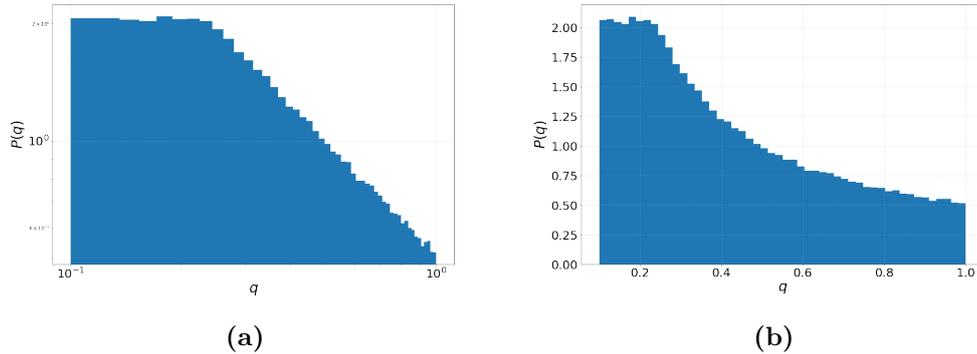


Figure 5.5. Probability distribution for the q parameter as described in [99]. The minimum secondary mass is set to $5M_{\odot}$. In panel (a), the distribution is double-log scale, while linear in panel (b).

To comprehend how cross-correlation influences the posterior, let's examine the scenario of an infinitely dense and wide, homogeneous distribution of hosts. In this case, the posterior distribution would be flat, since for each values of H_0 , there will always be a host with a z_{α} such that the exponent in eq. (4.46) is zero. The picture gets more complicated if the sample is not infinite, i.e. for some value of H_0 there are no z_{α} that are root of the exponent. In this case, the posterior will not be totally flat, but will have a maximum, although less informative than the non infinitely dense scenario.

Spatial correlation between DSs and hosts, along with auto-correlations of hosts, sharpens the peak of the posterior. This is because it enhances the probability of finding a likely host in a region of increased galaxy density for the correct value of H_0 . While this may not hold true for a single DS, combining multiple DS events results in a sharpening of the posterior's peak centred on the exact H_0 value.

This effect persists even when the galaxy host is not explicitly included in the host catalogue. There will likely be a galaxy in that sample spatially correlated with the missing host, contributing to the overall sharpening of the posterior distribution.

5.4 Estimating the Likelihood

The global likelihood of the DS distance given H_0 should be estimated by integrating the likelihood of each DS over a volume encompassing the full catalogue of hosts. Assuming Gaussian likelihood and negligible errors on the measured redshift of the host, the result is a sum of error functions centred on all objects in the host catalogue, as given by eq. 4.46. However, summing over millions of objects is computationally challenging and unnecessary. This is because the contribution of objects, which in fact constitute the majority, that are far from the DS location is exponentially suppressed.

Therefore, a more computationally efficient approach can be employed by focusing

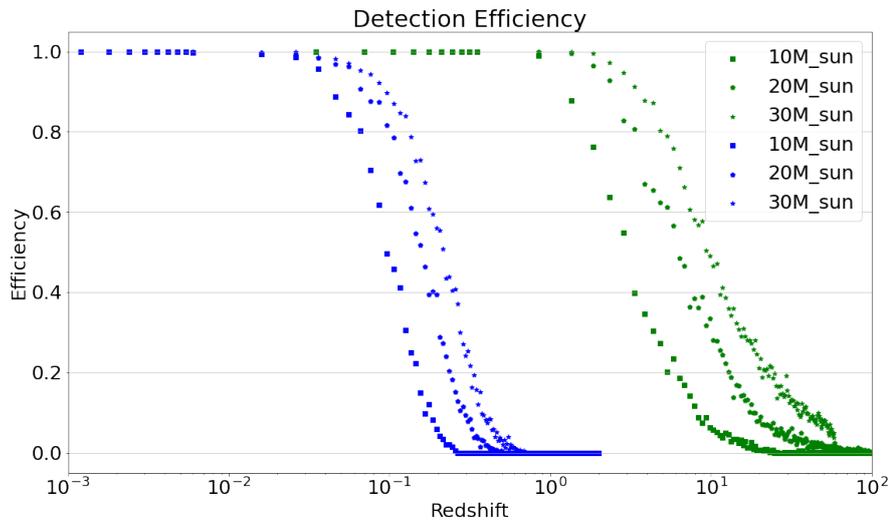


Figure 5.6. The efficiency for aLIGO and ET-D in blue and green respectively. The aLIGO network includes only the two LIGO interferometers with the advanced noise figure. A DS is detected if $\rho_{tot} \geq 8$. The mass refers to the total mass in the SFR.

on the relevant subset of host objects that are close to the DS location, where the contribution to the likelihood is significant, thus avoiding unnecessary computational burden. This is achieved by defining, for each DS, a "searching cone"³. Given the measured position of the DS, which does not depend on H_0 , the summation is restricted to those hosts located within a truncated cone centred on the DS. The cone's size is determined by requiring the angular coordinates and estimated luminosity distance (a H_0 -dependent quantity) to be within 3.5σ from the DS, where σ represents the uncertainty on the DS-measured angular coordinates and distance.

In other words, given the spatial coordinates of the i -th DS, (Dl_i, θ_i, ϕ_i) , measured with Gaussian uncertainties $(\sigma_{Dl}, \sigma_\theta, \sigma_\phi)$, the summation is performed only over those host galaxies located within the cone $(Dl_i \pm 3.5\sigma_{Dl}, \theta_i \pm 3.5\sigma_{\theta,i}, \phi_i \pm 3.5\sigma_{\phi,i})$. The dependence of the host's estimated distance $Dl(z, H_0)$ on the Hubble constant has the potential to extend the size of the searching cone beyond the boundaries of the host catalogue along the radial direction. In most applications and tests, the DSs sample is selected in such a way that the searching cone remains within the boundaries of the host catalogue for all H_0 values allowed by the prior.

It is worth noticing that, in the current implementation of the pipeline, all uncertainties are assumed to be Gaussian, whereas the angular error box of the GW events actually observed is more irregular. It will not be a difficult task to include a specific module in the pipeline that defines a 99 % probability area of specified shape around the DS event, which I plan to do for more realistic ET forecast once a stable design of the interferometer will be adopted.

³Also "search box" or "research box" or simply box.

In the likelihood expression, eq. (4.13), σ_i represents the uncertainty on the assumed DS distance $Dl(z, H_0)$, while in the following tests the DSs sample will be characterised by the relative errors in the measured distance. It is important to clarify that DSs are not placed on top of a host galaxy. Instead, a Gaussian scatter both in the radial and angular direction has been applied to determine the location of the DS around the host. This procedure mimics and accounts for the observational error in estimating the spatial location of the DS.

5.5 Testing the Pipeline: Robustness tests.

The purpose of this section is to validate the pipeline, and for this, several sanity checks and robustness tests have been designed and executed. The catalogues of galaxy hosts and DS were both extracted from the same uniform parent catalogue of unclustered objects.

The host catalogue is characterised by being both wide and deep, with $\phi \in [0, \pi/2]$, $\theta \in [0, \pi/2]$ and $z \in [0.05, 2.5]$, but intentionally sparse with $\mathcal{O}(3 \times 10^6)$ objects. The sparsity is chosen to alleviate computational burden, resulting in a posterior distribution that is noisy rather than smooth. It's important to note that the presence of Poisson noise in the posterior distribution does not impact the results of these tests.

Throughout these tests, the "searching cone" associated with each DS is carefully maintained within the boundaries of the host sample, ensuring a controlled experimental environment. The outcomes of these tests serve to verify the robustness and reliability of the analysis pipeline.

In all of this work, the prior on H_0 is always chosen flat between 55 Km/s/Mpc and 85 Km/s/Mpc for the robustness test. The prior is still assumed flat afterwards in Sec.5.6, with lower and upper boundaries of 60 Km/s/Mpc and 76 Km/s/Mpc respectively.

5.5.1 Robustness to sample selection effects

Galaxy host catalogues, including the Euclid spectroscopic one, are inherently incomplete due to selection effects introduced during the observation process. Most galaxy samples are flux-limited, meaning they only include objects whose apparent luminosity is above the detection threshold. This selection criterion alters the number density of objects, causing it to decrease with the distance (and redshift) of the sources, resulting in a redshift distribution $N(z)$ different from the intrinsic one.

This modulation has the potential to affect Bayesian inference by introducing a spurious z -dependence in the likelihood analysis. As illustrated in the previous chapter, to correct for this effect, a set of z -dependent statistical weights $\omega(z)$ is introduced to restore a constant number density in the host sample.

While these weights effectively eliminate the potential bias induced by observational selection, it's important to note that they do not reduce the Poisson noise, which is amplified by having selected a subset of potential hosts.

To assess the effect and confirm that the introduction of these weights effectively eliminates bias, I conducted a simple robustness test. This involved comparing the results of two analyses using the same set of DSs, both extracted from a uniformly distributed parent sample.

In the first case, the host catalogue is also uniformly distributed with a constant comoving density, mimicking the scenario of an ideal, volume-limited sample. In the second case, a subsample of hosts is considered. It is also uniformly distributed but has a z -modulated number density, reflecting a non-monotonic $N(z)$ characteristic of a flux-limited galaxy sample.

In this second case, the statistical weights are computed in z -bins as the ratio between the number of galaxies in the flux-limited and volume-limited catalogues. This comparison helps evaluate the impact of observational selection effects and the effectiveness of the introduced weights in mitigating bias.

The results, obtained for the same catalogue of 150 mock Ds, are shown in Fig. 5.7. The figure shows how the two results are not only compatible with each other but also with the true value of H_0 . This finding confirms that the correction introduced does not induce a bias in the posterior, and that the weight strategy is correct and properly implemented. Before proceeding with the rest of the presentation, it

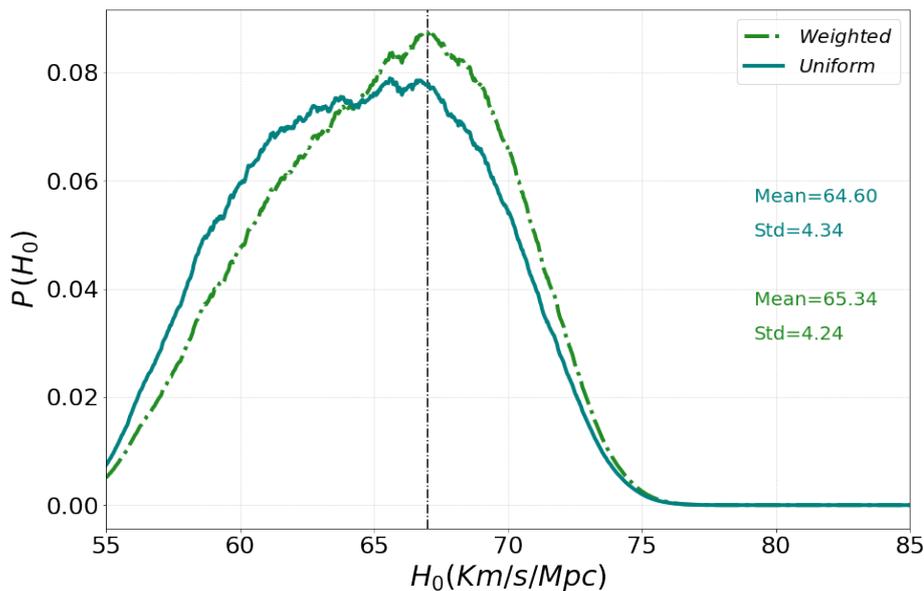


Figure 5.7. The posterior for the cross-correlation with CODE of an ensemble of 150 DSs correlated with a uniform Universe, teal curve, and in green the same set of DSs cross-correlated with a catalogue extracted from the uniform one with a given $N(z)$.

is useful to discuss Fig. 5.7, especially in light of the subsequent "Uniform-Wide" implementations. As shown, for example, in Fig. 5.8, the uniform implementation is less informative compared to the weighted one, which a priori represents a subsample. Although the results are compatible within one sigma, it is appropriate to comment

on this outcome.

As evident from the comparison with other subsequent uniform implementations, represented by the blue curves, uniform runs generally do not exhibit bias. However, a more detailed analysis has revealed that, due to mere statistical fluctuation, the DS have been scattered in such a way that the mode of the sample is higher than the mode of the non-scattered sample. By inverting the Hubble's law, we understand that this favours lower values of the Hubble constant compared to those under examination.

5.5.2 Robustness to the choice of the galaxy host sample

Ideally, the results of the analysis should not depend on the relative size and possible overlap between the sample of hosts and that of DS. To assess this, a series of tests have been conducted. In the first test, which is outlined in this section, the DS catalogue is kept fixed, and the volume of the host sample is varied.

The reference DS catalogue includes 150 objects uniformly distributed over a sky octant. It lists the luminosity distance and the angular position of each event. Distances are assumed to be measured with 10% relative error. Angular positions have the same uncertainty quantified by a 2D Gaussian probability with a radius such that $\Omega_{90} = 10 \text{ deg}^2$. A reference 'true' value of $H_0 = 67 \text{ Km/s/Mpc}$ is assumed to perform the analysis. The DSs span a Dl range between 2257.3 Mpc to 9875.11 Mpc.

To conduct the test, I have utilised various samples of host galaxies. These hosts are uniformly distributed over the same octant as the DS but span different redshift intervals, with each interval becoming progressively smaller in size. This variation in the size of the redshift intervals allows for a systematic examination of the impact of the DS sample's volume on the analysis results. The largest sample, used as a reference and referred to as *Uniform-Wide*, spans the redshift range $0.05 \leq z \leq 2.5$. This reference sample serves as a baseline for comparison. The resulting posterior probability distribution for H_0 obtained in this case is shown in Figure 5.8 with a continuous teal curve.

Following that, I systematically reduced the redshift range of the host sample, resulting in the creation of the *Uniform-CUT00* catalogue ($0.1 \leq z \leq 2$, green continuous curve), the *Uniform-CUT01* catalogue ($0.2 \leq z \leq 1.8$, orange dashed line), and the *Uniform-CUT02* catalogue ($0.3 \leq z \leq 1.6$, dotted dark-red). The posterior distributions obtained from the analysis of the first three samples (*Uniform-CUT00*, *Uniform-CUT01*, and *Uniform-CUT02*) are almost superimposed on each other and peak in correspondence to the expected H_0 value. However, when the smallest host catalogue is used in the analysis, the peak of the posterior is slightly offset, resulting in a small underestimate of H_0 . While in principle, if the host catalogue fully covers the research cone, there is no reason to think that there will be a bias, things have to be proven, and we make this test more to check the health of the code.

The observed mismatch, while statistically limited in significance, may be attributed to the fact that, in this case, for certain choices of H_0 , the searching cone of some DS hosts is not fully contained within the volume of the host sample. Further

investigations seem to support this. Since the code was then updated and the deviation is not so significant, the test still proves the point⁴.

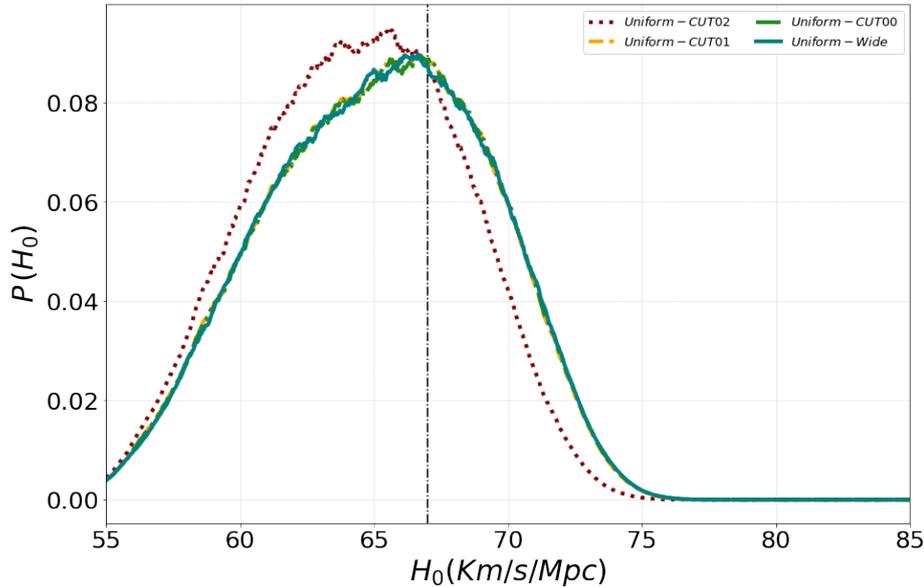


Figure 5.8. All the posteriors for the fixed DSs in the distance range between 2257.3 Mpc and 9875.11 Mpc. In teal the posterior for the *Uniform-Wide* hosts catalogue, in green, the distribution for the *Uniform-CUT00* catalogue, the orange dashed line refers to the posterior in the *Uniform-CUT01* case and the *Uniform-CUT02* results are shown via the dotted dark-red curve. Only the *Uniform-CUT02* case differ from the others.

5.5.3 Robustness to the choice of the DS sample

In this new test, the DS sample is changed while maintaining the *Uniform-Wide* sample as the host galaxy catalogue for each run. The reference DS sample used in this test, spanning a range in luminosity distance equal to (6827.4; 15978.6) Mpc, is the same as in the previous section, resulting in the same posterior curve (teal continuous curve in Fig 5.9 as the one seen in Fig. 5.8).

We considered three more DS samples. They all share the same sky area and have the same distance and angular position errors. What changes is the range of radial distance spanned by the objects bin the catalogues, namely: (2257.3, 9875.1) Mpc for the sample dubbed *DS-CUT00*, (3660.3, 11239.4) Mpc for the sample dubbed *DS-CUT01* and (6827.4, 15978.6) for the sample dubbed *DS-CUT01*. The number of DS in not the same in all catalogues. The number density is.

The results depicted in Fig 5.9 indicate that the four posterior distributions agree with each other. The most discrepant curve is that of the *DS-CUT00* sample. However, its displacement with respect to the reference case is hardly significant.

⁴The code changed after this run, with a more strict control over the selection of DSs.

The difference in the position of the peaks of the curves falls well within the standard deviation of these curves, $\sigma \sim 3.40$ Km/s/Mpc, as estimated from their variance. Since this test was performed after some improvement of CODE, the search cone is now always contained in the host catalogue.

As expected, no significant dependence of H_0 on the spatial location of the DS sample

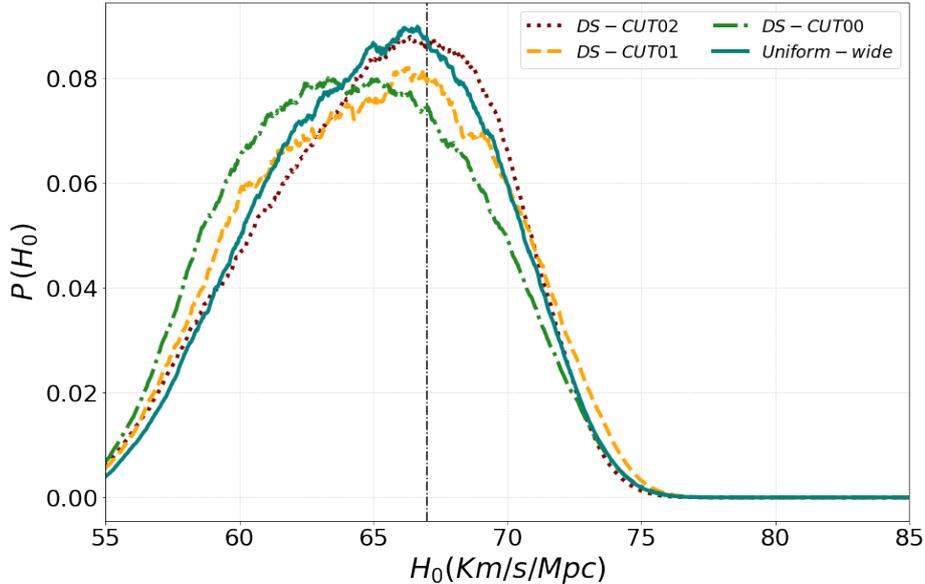


Figure 5.9. All posteriors for the fixed host case, named after the DSs catalogues. In teal, the reference curve, the same as in Fig. 5.8. The dot-dashed green line for the *DS-CUT00* case. In orange, the posterior for the catalogue *DS-CUT01*, and finally in dark red the *DS-CUT02* curve. All the curves are compatible with the reference value $H_0 = 67$ Km/s/Mpc. The curves are compatible within each other.

5.5.4 Robustness to the redshift distribution of the host galaxies

The focus now shifts to changing the host selection function $N(z)$. These tests not only provide insights into the robustness when varying the distribution of the hosts but also serve as further evidence that the strategy of statistical weights is effective.

A new run with the *Uniform-Wide* catalogue is used as a reference, with a change in the sample of DSs. It is important to note that with the introduction of the selection function, the presence of the galaxy associated with the i -th DS in the host catalogue is no longer certain since the extracted catalogue possesses $\sim 33\%$ of *Uniform-Wide* hosts.

To obtain the new catalogue, a Monte Carlo rejection is performed. Given the chosen $N(z)$, the total number of hosts is set and an ensemble of redshifts are drawn following the selection function. The next step is then to bin those ensembles and

then, for each bin, extract from *Uniform-Wide* N_i hosts, where N_i is the absolute number of hosts in the i -th bin. It was ensured that each bin of *Uniform-Wide* contains more objects than the same bin for the selected $N(z)$, thus avoiding double extractions due to missing possible entries.

The two selected selection functions resemble the functional form of the flagship, expressed as

$$N(z) = z^\alpha e^{-\frac{z^\beta}{\gamma}}. \quad (5.4)$$

Two $N(z)$ are tested: Nz00 and Nz01, defined by $\alpha = 2, \beta = 1, \gamma = 0.5$ and $\alpha = 2, \beta = 1.82, \gamma = 1$. Both sets of parameters are chosen so that the respective $N(z)$ peaks between $z = 0.9$ and $z = 1.1$, as visible in Fig. 5.10 to mimic the expected Euclid spectroscopic survey behaviour. The test indicates that CODE is not overly

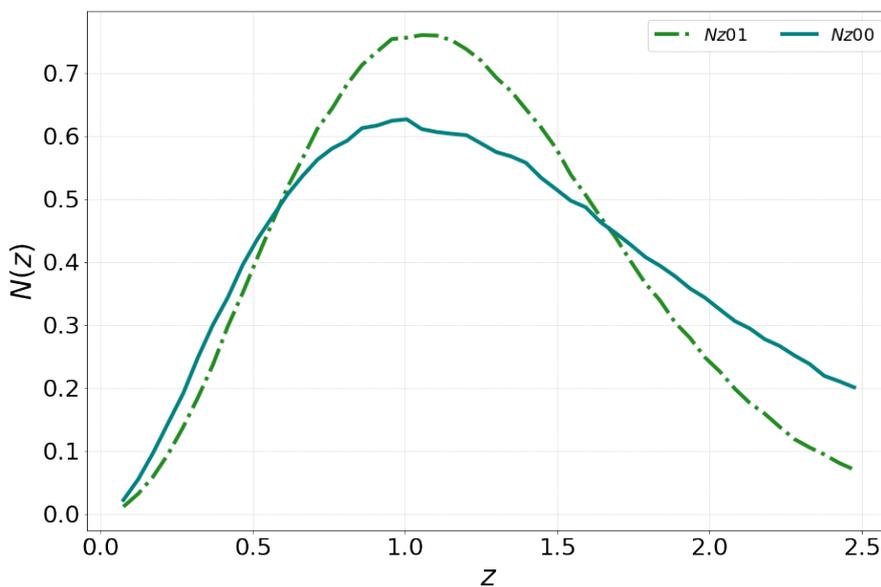


Figure 5.10. Comparison between the two selection functions, in teal Nz00 and in green Nz01. Both curves are normalised so that the total area is 1.

sensitive to the selection function of the hosts, see Fig. 5.11. Due to selection effects, the actual host galaxy of the DS may not be included in the galaxy catalogue. However, as highlighted in several studies [109, 6, 103], the presence of the host galaxy in the catalogue is not essential for effective Bayesian analysis, provided that the selection function of the catalogue is known and there is a physical correlation between the galaxies and the missing host. In fact, if the galaxies in the catalogue were not physically correlated with the DS, the uncertainty in the estimation of the Hubble constant would be influenced by two competing factors: the mean separation between galaxies and the uncertainty in the distance of the DS. If the former were smaller than the latter, the DS could be associated with any galaxy within the DS uncertainty volume, leading to an error in the measured DS distance propagating into

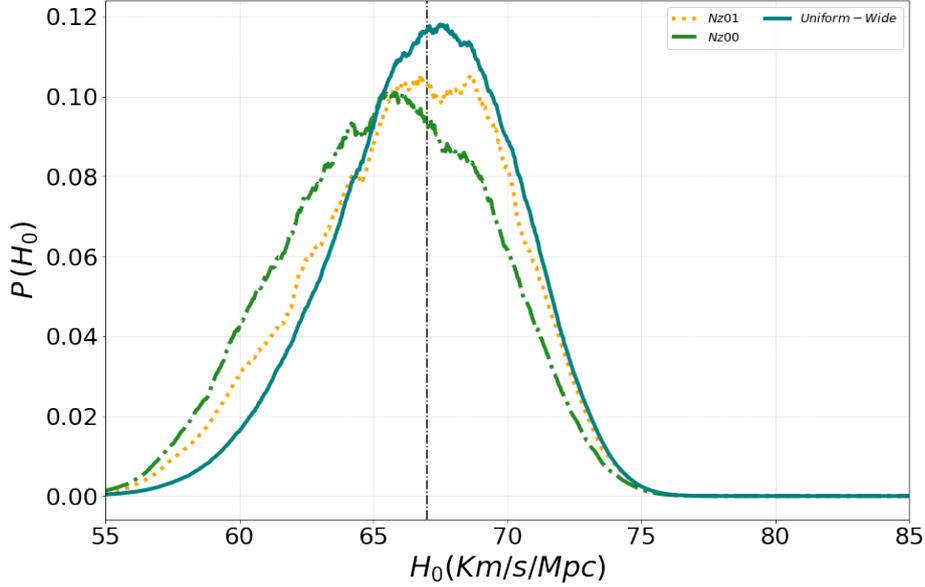


Figure 5.11. This figure shows all the posteriors for the $N(z)$ robustness test. In teal, the reference for the *Uniform-Wide* case. The dotted orange line refers to the Nz01 catalogue, while the green dot-dashed curve refers to the nz00 case. Both curves are compatible with the reference and have standard deviations of $\sigma_{00} = 3.74$ and $\sigma_{01} = 3.65$ for Nz00 and Nz01 respectively.

that of H_0 . Conversely, if the error in the DS distance were smaller than the mean galaxy separation, the DS might be associated with galaxies outside the uncertainty volume, resulting in a systematic error in H_0 . The existence of a physical correlation between the missing host and the observed galaxies helps mitigate both types of errors by increasing the number of observed DSs.

While the run with the Nz00 selection function may seem slightly off, the standard deviations of the posteriors are similar with numerical values $\sigma_{00} = 3.74$ and $\sigma_{01} = 3.65$ for Nz00 and Nz01 respectively. The explanation for the lower standard deviation reads as follows. The DSs are common to both runs and have been extracted from the *Uniform-Wide* catalogue. Since the *Uniform-Wide* catalogue has an increasing number of possible DSs with distance, the selection function with the largest peak in z will have a higher chance of containing the true host. This is because, in that region, it will be more populated and therefore, a higher probability of containing the right host. This explains the slight improvement in the standard deviation of run nz01.

It must be addressed why in this realisation the teal curve, the uniform case, is more informative than in the previous runs. Again is a matter of statistic, that will be completely removed once we move to the forecast section since the number of DSs increase. If we consider the difference between the scattered distance of the DS and the true luminosity distance, we can then look at the dispersion of the scattered

sources. For this realisation, the sources were scattered particularly close to the real position. To give some numbers, no source was scattered more than 2σ apart from the real position, resulting in a particularly informative run.

5.5.5 Robustness to the radial distribution of the DS

This test aims to determine whether a non-uniform distribution of sources, hence modulated by a certain distribution function $N(Dl)$, can influence the determination of H_0 . The reason why, at least in our case, a dependence on the source distribution is not expected is that the selection process remains Poissonian. Given a host distribution, a DS is randomly associated with a host. If it is assumed that the DSs are now distributed according to an $N(Dl)$ and then sampled, still using a Monte Carlo, the only effect will be to have more sources for a certain range of distances. This does not introduce a bias because the likelihood for H_0 , $\mathcal{L}(H_0|Dl)$ remains the same. Correction for any deviations from a uniform distribution must therefore be compensated at the host level, as described in the previous section, through the introduction of the statistical weights $\omega(z)$.

This sampling method avoids the problem of "double counting" addressed in [103]. The issue arises when sources are extracted from a host sample with a certain selection function, themselves with their distribution function, and not through sampling. For example, if a sample of galaxies has a $N(z) \sim 1/z^2$ and a DS host assignment probability $P \sim 1/z^2$ is imposed, the resulting source population will have a distribution of the type z^4 . If, instead, the initial host population is directly sampled, the DSs will follow the same distribution.

The selected $N(Dl)$ is defined over the region $Dl \in [Dl_{min}, Dl_{max}]$ with $Dl_{min} = 6800$ Mpc and $Dl_{max} = 11000$ Mpc as follows,

$$N(Dl) = \left(\frac{Dl}{p}\right)^\alpha e^{\left(\frac{Dl}{p}\right)^\beta}, \quad (5.5)$$

with $\alpha = \beta = 2$ and the pivot $p = (Dl_{max} + Dl_{min})/2$.

The test is then conducted with the following setup: a control run on the *Uniform-Wide* catalogue, again with 150 DSs extracted from the uniform catalogue. A sub-catalogue of DSs is then extracted from the uniform one, having the distribution function $N(Dl)$ defined in eq. (5.5). The results of the analysis are visible in Fig. 5.12. In teal, the uniform run can be seen, while the dot-dashed green curve refers to the run in which the DSs are extracted according to $N(Dl)$. As can be seen in Fig. 5.12, the two posteriors are very similar, even in terms of standard deviation, which is $\sigma \sim 4$ Km/s/Mpc. This test has thus demonstrated two important properties. The first is that the DSs selection method implemented does not introduce bias and avoids the problem of double counting. The second property demonstrated is that it will be possible to sample DSs from the flagship without having to introduce any compensation.

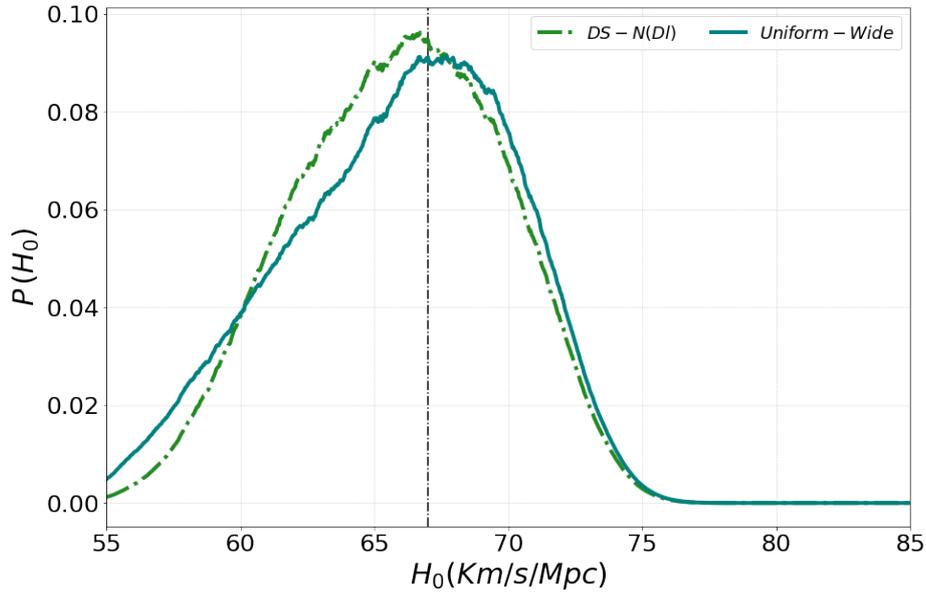


Figure 5.12. Comparison between posteriors for the run with the DSs following the $N(DI)$ in eq. (5.5) in green and a reference case in teal. The number of DSs is in both case 150 and the hosts catalogue is always *Uniform-Wide*.

5.6 Forecasting

In this section, I apply the validated pipeline, as demonstrated in the tests presented in the previous sections and Chapter 4, to a more realistic scenario involving a DS catalogue from a GW observational campaign conducted with the next-generation interferometer ET. This is correlated with a catalogue of potential galaxy hosts designed to mimic that of the Euclid spectroscopic redshift survey.

While the primary goal of this section is to operate the CODE pipeline on a more realistic dataset to showcase its potential for future applications, I will also address the question: "How well will we be able to estimate H_0 from a large sample of dark structures at moderate redshift, taking advantage of the availability of wide and deep galaxy surveys covering an unprecedented volume of our Universe?"

I will approach this problem by steps of increasing complexity using increasingly realistic datasets.

The simplest one is an ideal sample, referred to as *Uniform-ideal*. This catalogue consists of unclustered objects with a constant number density equal to that of the galaxies in the redshift bin $z = [0.9; 1.1]$ in the simulated Flagship spectroscopic galaxy catalogue. It covers an octant of the sky and a redshift range $z = [0.79, 2]$ containing $7 \cdot 10^6$ objects.

The second, more realistic sample, also consists of a set of unclustered objects distributed over the same octant and over the same redshift range but with a radial distribution matching the $N(z)$ measured in the Flagship spectroscopic galaxy

catalogue. This catalogue is dubbed Uniform-Flagship. Finally, I will also consider the Flagship true catalogue itself, containing a set of clustered, $H - \alpha$ line emitting objects above the expected Euclid detection threshold. I will refer to this catalogue as True-flagship.

Several sets of DSs will be considered along with the host catalogues. The reference simulated dark sirens catalogue consists of a collection of ET "golden events" characterised by objects whose distance has been estimated with a 10% uncertainty and angular coordinates determined within a 90% uncertainty area of $\Omega_{90} = 10 \text{ deg}^2$. Both uncertainties are assumed to be Gaussian.

The location of these DS has been determined by randomly choosing a host galaxy in the host parent sample, which can either be the Uniform-ideal sample or the spectroscopic flagship catalogue itself, and then applying Gaussian random scatter, both in the radial direction and on the sky, to the host.

For this reason, when considering the Uniform-Flagship host catalogue, the DS used in the likelihood analyses may not have their true host in the sample. This highlights the inherent uncertainty in associating DS with specific host galaxies, especially when using a galaxy catalogue that may not include all potential hosts.

5.6.1 Sensitivity to the DS angular position uncertainties

Reducing the uncertainty in the angular position of the dark sirens, Ω_{90} , is expected to narrow the posterior on H_0 [81]. The sensitivity of the H_0 estimate to the uncertainty Ω_{90} is indeed large for nearby DS, where the number of potential hosts within the uncertainty box is relatively small. However, for DS located at moderate redshifts, such as those considered here, the volume of the uncertainty box is large and contains a correspondingly larger number of potential hosts. As a result, the sensitivity to Ω_{90} is expected to be weaker.

To test this hypothesis, I have generated a sample of 300 DS extracted from the Uniform-ideal catalogue, "observed" with four different angular resolutions: $\Omega_{90} = 5 \text{ deg}^2$, 10 deg^2 , and 20 deg^2 . their parent hosts are located well within the sky octant to ensure that the 10σ angular uncertainty box associated with each dark siren does not extend beyond the octant's boundaries. The results are shown in Fig. 5.13.

The result of this test, shown in Fig. 5.13, demonstrate that the width of the posterior increases when the uncertainty in the DS angular position increases, as expected. The magnitude of the effect, however, is smaller than the one measured in the Bayesian analysis of simulated data mimicking LIGO-Virgo-KAGRA O4 and O5 observations of [9]. The reason of the mismatch is probably due to the fact that, in my analysis, I have ignored any correlation between the relative uncertainty of the luminosity distance and the sky location area of the DS. Such correlation is expected to be present, since both uncertainties correlate with the SNR of the event, as proven in figure 3 of [9]. Although in [9] the correlation is clear, one can see that in figure 3 of [9], for the range of uncertainties explored in this work, $\sigma_{Dl} \sim 5 - 10\%$ the correlation can be omitted, since the spread of the points. It is true that this approximation must be overcome to obtain more realistic and robust results, as done

in codes already mentioned. For a golden DSs sample obtained from an ET-like observational campaign, the correlation tightens and should not be ignored and will be explicitly included in the next release of my pipeline. Having ignored correlations, the number of potential hosts in the tests of Figure Fig. 5.13 increases linearly with the size of the sky location area, being ~ 11000 for $\Omega_{90} = 5 \text{ deg}^2$.

This high average number of hosts is due to the fact that we are moving to medium-high redshifts, and that if we assume that galaxies are uniformly distributed in volume, then the number of hosts grows as the volume. A more in-depth study taking into account the correlation between the variables at play at these redshifts is underway.

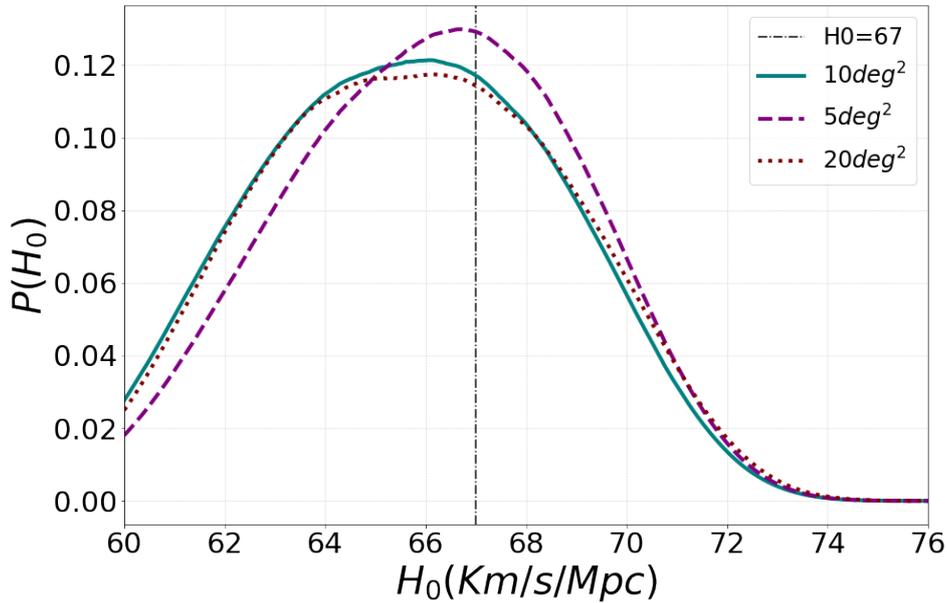


Figure 5.13. This picture shows all the posterior for the tested Ω_{90} . In teal, the reference run with $\Omega_{90} = 10 \text{ deg}^2$, the dash-dotted green line shows the posterior for the $\Omega_{90} = 5 \text{ deg}^2$ case while the orange curves represent the results for $\Omega_{90} = 20 \text{ deg}^2$. The DSs are drawn from the Uniform-ideal catalogue.

5.6.2 Sensitivity to the uncertainty in the DS Luminosity Distance

Having confirmed that the results are weakly insensitive to the error in the angular position of the DS, as long as these are kept within the range $\Omega_{90} = [5, 20] \text{ deg}^2$, the focus is shifted toward the errors in the luminosity distance, σ_{DL} .

We anticipate that the impact of σ_{DL} on the precision in the estimate of H_0 is more significant than that of Ω_{90} . This is because doubling the error in the angular position increases the size of the error box by a factor of two, whereas doubling the error on the luminosity distance increases the volume of the error box by a larger factor, depending on the size of the error.

One issue arises when performing this test. The first one concerns the radial boundaries of the host sample. The size of the searching box associated with a DS depends on the assumed value of H_0 and on the amplitude of σ_{DL} . For a given H_0 value, only DS within the observed luminosity distance range $D_L(z_{min}, H_0), D_L(z_{max}, H_0)$, where z_{min} and z_{max} are the minimum and maximum redshifts of the galaxies in the host catalogue, are located within the volume of the galaxy sample.

Since the likelihood analysis spans a range of H_0 values set by a flat prior on this parameter, only a limited number of DS are guaranteed to be in the host volume for all the H_0 values considered. The larger the error on the observed distance, the smaller the volume of the DS sample fully embedded in the host catalogue. There will be a value of σ_{DL} above which no DS sample is fully contained in the host galaxy catalogue for all the H_0 values.

It turns out that for relative errors of $\sigma_{DL} = 20\%$ of the observed luminosity distance, no DS sample is fully contained within the mock galaxy catalogue. As a result, we need to correct for the boundary effects in the likelihood analysis. This involves accounting for the fact that, above or below certain H_0 values, DS may fall outside the boundaries of the galaxy sample.

To address this issue, we need to implement a correction procedure in the likelihood analysis that, as described in the previous chapter, consist in properly modelling the $\beta(H_0)$ normalisation term. The new β function is the one proposed in eq. (4.39) and the definitions in eq. (4.40) and eq. (4.41). When the research area is fully embedded, beta is described in eq. (4.28). The first step is to determine the extent to which the borders can be exceeded. A preliminary test is conducted with $\sigma_{DL} = 20\%$ to check for any bias. It is noted that to perform these tests, the samples of DSs must differ, as different values of σ_{DL} require different scattering. A set of 300 DSs, extracted from the Uniform-ideal catalogue and cross-correlated with the Uniform-Flagship one, is now tested.

As demonstrated by Fig. 5.14, there is no bias on H_0 , indicating that the correction to beta does not introduce such an effect. However, the belief that the correction is not normalising the posterior correctly persists. Although it is true that the two runs involve different sets of DSs, and the 20% run might simply have been a 'lucky' one, this conclusion is not entirely convincing. Additional realisations at 20% have been conducted: none show a systematic bias, but all are similar in height, thus compatible with the 10% run. Given the unlikelihood of each realisation of the DSs having particularly small scattering despite the greater standard deviation, it is concluded that the method implemented is not sufficient to correct for edge effects once such uncertainties are reached.

For cross-correlation purposes, this limitation is not overly significant, considering the number of gold events expected to be detected by ET. However, it is deemed appropriate to further investigate this aspect, in a manner similar to that addressed in [103], despite the differences in the physical interpretation of the quantities. Correcting for edge effects opens the possibility of cross-correlating using the entire host catalogue, thereby increasing the number of DSs admitted in the analysis, especially in high-redshift catalogues, where a slight increase in distance implies a

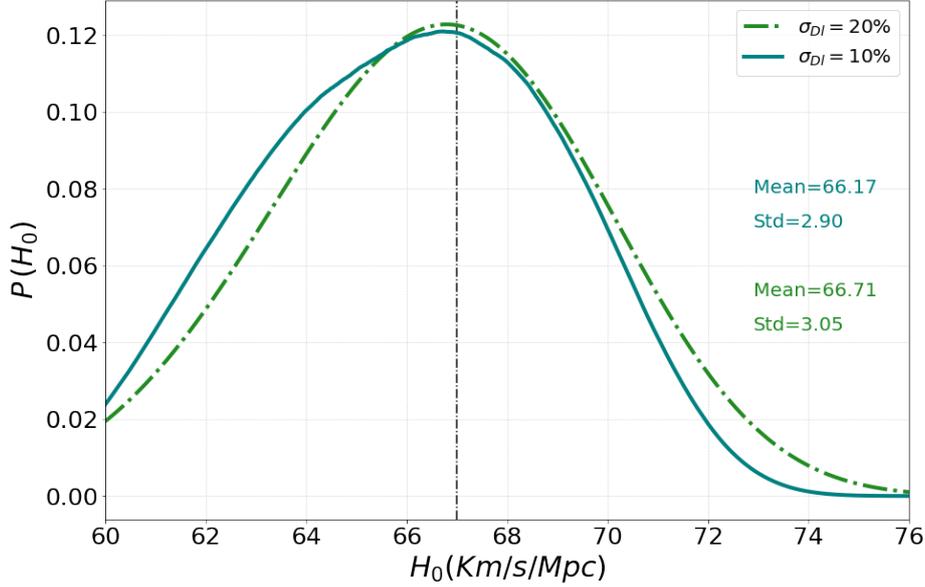


Figure 5.14. The reference posterior in teal and the $\sigma_{DI} = 20\%$ case in green. Both runs have 300 DSs cross correlated over the Uniform-Flagship catalogue.

large increase in volume.

Subsequently, it is verified that for a 15% uncertainty, the modification to the β function yields the expected results. As observed in Fig. 5.15, the correction appears to be functioning effectively. The curve with the greater uncertainty is wider compared to the reference one, and as definitive proof, a second run with $\sigma_{DI} = 15\%$ without the correction was launched for comparison. Fig. 5.16 confirms the improvement of the posterior with the correction outlined in chapter four, at least for this category of events.

The analysis then continues to examine how the posterior changes with varying σ_{DI} values at 1%, 5%, 10% (the reference), and 15%. As demonstrated by Fig. 5.17, σ_{DI} is the parameter that predominantly governs the quality of the posterior. The precision leap in the 1% run is such that it renders the other posteriors off-scale. To better appreciate the variation between 10% and 5%, the run at 1% is omitted in the subsequent plot. It is observed that the increase in precision is not linear. While there is a gain in precision between the 15% and 10% cases, reducing the uncertainty by another 5% leads to sufficiently narrow posteriors after 300 events. It should be noted, as shown in Fig(3.7), that within the Euclid survey footprint, ET expects between 1000 and 2000 events in 5 years of observations, making the 5% case not just a Utopian idea, but a concrete reality.

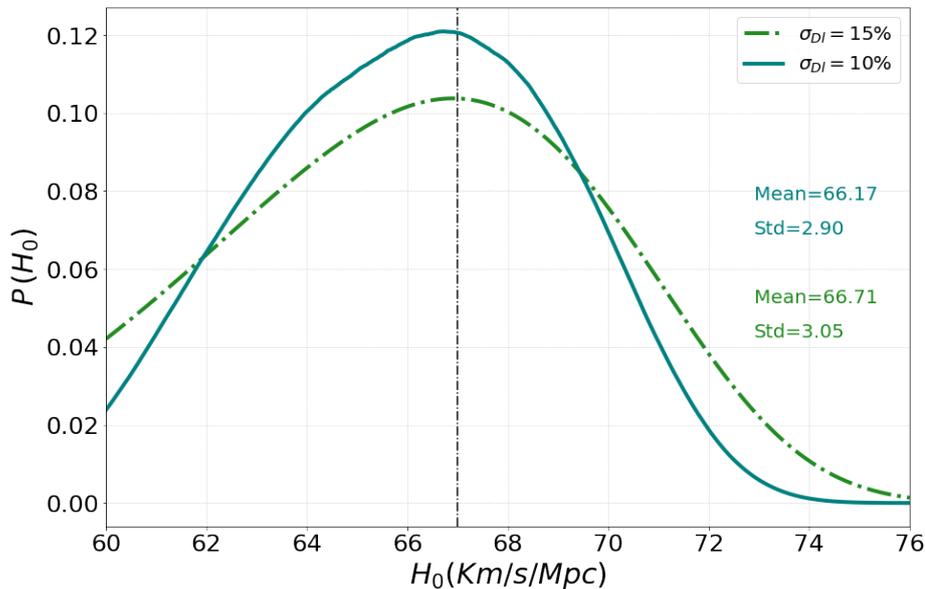


Figure 5.15. The reference posterior in teal and the $\sigma_{DI} = 15\%$ case in green. Both runs involve 300 DSs cross-correlated over the Uniform-Flagship catalogue.

5.6.3 Dependence on the number of dark sirens.

When dealing with transient events like dark sirens, one of the critical questions to address is how many events have to be observed to achieve a target accuracy on the relevant physical quantity one wants to measure. This requirement ultimately sets the lifetime of the experiment, considering factors such as the event rate, detector sensitivity, and duty cycle of the apparatus.

Our target quantity in this case is H_0 , the Hubble constant. Therefore, in this section, I will address a closely related question: is there a well-defined scaling of the H_0 uncertainty, σ_{H_0} , with the number of detected DS, N ? To investigate this, I consider "golden events," of which I am considering 4000, and perform the likelihood analysis using progressively larger sets of DS. I start from the first 100 sources, adding 100 more in each subsequent analysis. All the DS are contained in a volume fully embedded within that of the galaxy host, and both are extracted from the Uniform Flagship catalogue.

The results are shown in Fig. 5.19, where the coloured curves represent the evolution of the posterior distribution as the number of DS increases. Starting from the flatter, blue curve obtained with the poorest DS samples, the curves become more and more peaked (green to yellow to red), ending up with a quasi-Gaussian shape (darkest shade of red).

The width of the curves, and thus the uncertainty on H_0 , decreases with N , as expected. Since the DS are selected in the same volume, there is no sample variance, and the only contribution to the error budget is shot noise. This shot noise, because

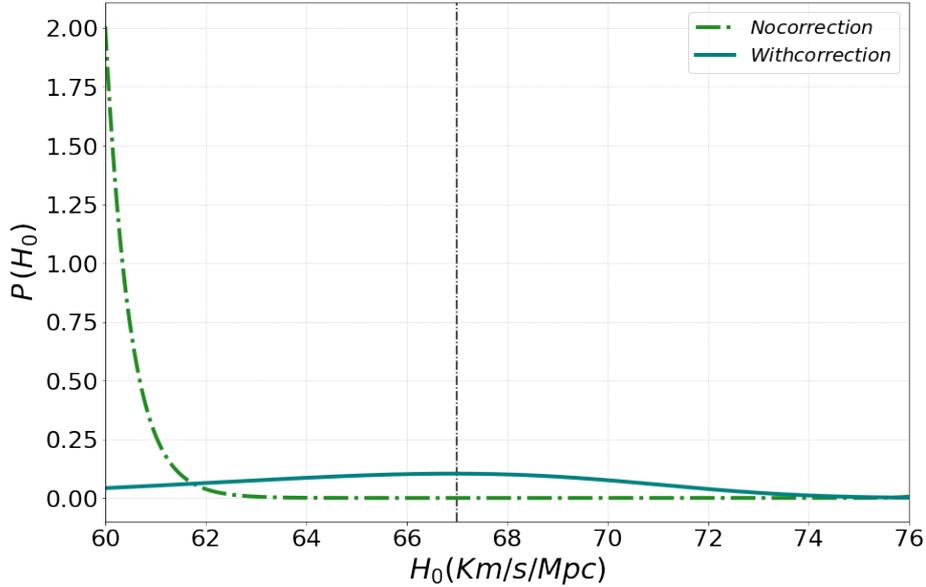


Figure 5.16. The corrected posterior in teal and the $\sigma_{DI} = 15\%$ uncorrected case in green. Both runs involve 300 DSs cross-correlated over the Uniform-Flagship catalogue. The teal posterior is similar to the green one in Fig. 5.15 but appears flat due to scale.

of the uniformity of the sample, should be purely Poissonian.

To verify this hypothesis, I plot in Fig. 5.20 the variance of each posterior curve in Fig. 5.19 as a function of the number of events (teal dots) and compare it with the expected Poisson noise $\sigma_P = H_0/\sqrt{N}$.

The expected Poisson scaling is indeed found in the large N limit (the curve that best fits the data for $N > 3000$ is $\propto N^{-0.44}$) whereas for small DS samples the dependence is steeper, reflecting the fact that the corresponding posterior curves are highly non-Gaussian for small N values.

To check if the accuracy in the estimate of H_0 depends on N Fig. 5.21. These results indicate that a 2% precision, comparable with that obtained with the best available H_0 probes, can be achieved by observing approximately 1500 dark sirens once we correlate them with the spatial positions of galaxies in a Euclid-like survey. One may suspect that using an ideal sample has led to overly optimistic predictions. However, I argue that this may not be the case, as in this test I am ignoring the fact that both dark sirens and hosts correlate with the underlying mass density distribution, and for this reason, are auto- and cross-correlated with each other.

I will show in the next section that the presence of this correlation further reduces the H_0 uncertainties, providing additional evidence for the effectiveness of the proposed methodology.

Another aspect to investigate is the relationship between the number of sources and the width of the posterior. For this purpose, a run with a deliberately inflated number of sources is conducted. Given the large pool of potential DSs, each event is

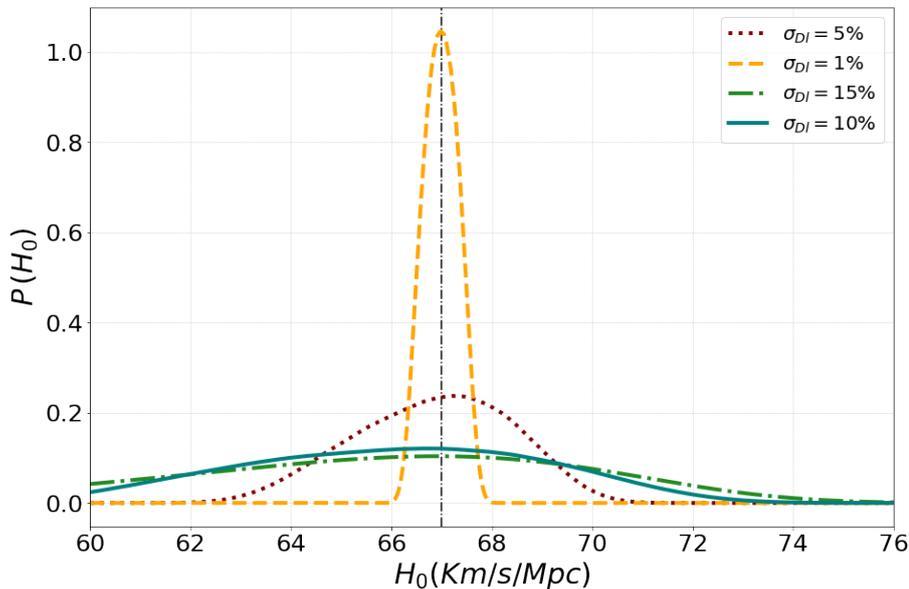


Figure 5.17. The collection of all posteriors for the σ_{Dl} runs. The number of DSs is 300 and the host catalogue is Uniform-Flagship

considered isolated and uncorrelated with others, aligning with the initial assumption of the formalism in use.

A run is performed with a number N of DSs equal to 4000. Due to the independence of individual events, it is possible to divide this run into a series of subsets of events S_n , such that if the set S_2 has N_2 events with $N_2 > N_1$, then all events present in the set S_1 will be present in S_2 and so forth.

For this test, events with $\sigma_{DL} = 10\%$, and $\Omega_{90} = 10 \text{ deg}^2$ are used as sample events. The events are arranged such that the research box is entirely contained in the catalogue, for every value of H_0 tested, and are cross-correlated with the Uniform-Flagship catalogue. The result of this test is depicted in Fig. 5.19.

Considering that independent events are being studied, it is reasonable to hypothesise that the uncertainty evolves proportionally to $1/\sqrt{N}$. Although the standard deviation in Fig. 5.20 decreases as $1/N^\alpha$, with $0 < \alpha < 1$, this observation does not align with the initial hypothesis. This discrepancy suggests that either the events are not entirely independent, or that, since in H_0 the curves are not purely Gaussian, the $1/\sqrt{N}$ behaviour only applies when the posteriors are narrow and their non-Gaussianity becomes negligible.

Concerning the independence of the events, an initial study in [103] concluded that the effect of non-independence plays a minor role in the magnitude of the uncertainty on H_0 . Therefore, the deviation from the expected behaviour is more likely attributable to non-Gaussianity. In principle, the posterior is not a Gaussian curve in H_0 , but for small σ , the posterior can be approximated as Gaussian. By

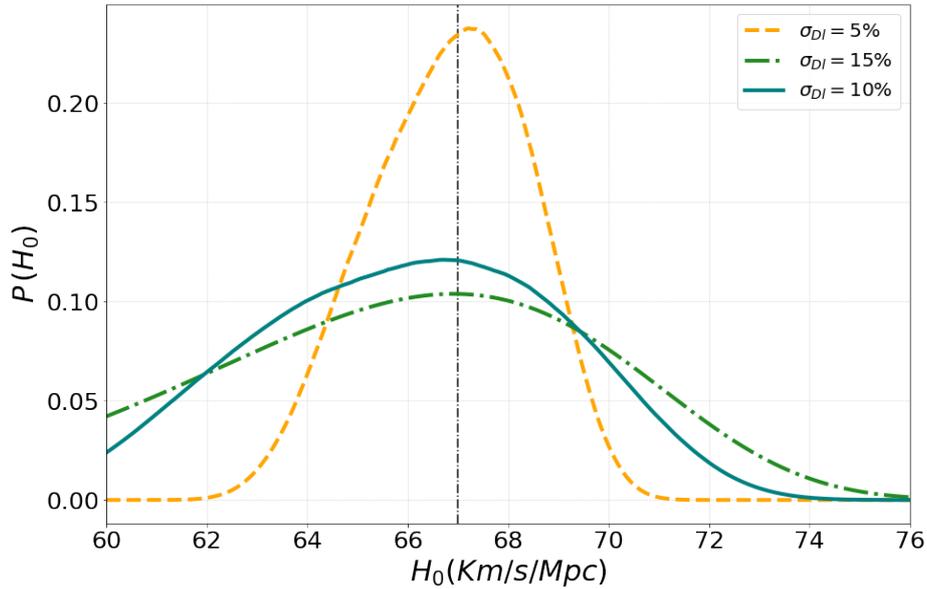


Figure 5.18. The collection of all posteriors for the σ_{Dl} runs. The number of DSs is 300 and the host catalogue is Uniform-Flagship. The run at 1% is omitted to better appreciate the variation between 10% and 5%.

selecting only the runs with $N > 3000$ and performing a fit of the type $y = 67/x^m$, a value of $m \sim 0.47$ is obtained, approaching the expected value of 0.5. This is also visible since from the tail of the σ Vs N curve in Fig. 5.20, since they almost lay on the curve.

The analysis now continues, showing in Fig. 5.21, how the mean, μ , of the posterior evolves with N . The figure clearly illustrates how the information builds up and the uncertainty reduce with an increasing number of DSs.

Given the shape of the posterior of each subset in Fig. 5.19, it is clear that, at least once the posterior is more bell shaped, the peak of the distribution and the mean are close, so the evolution of the peak is not showed.

In all these tests, and in what follows, it is always ensured that the research area is well contained within the catalogue to avoid this type of bias. This constraint is not an obstacle when calibrating a pipeline with synthetic data; however, in a real-world application, it may be detrimental as it might result in the loss of potential DS. Such an effect can be corrected, as explored in Sec. 4.1.3; however, at least for our implementation, the correction still has limitations. This type of correction is reminiscent of the so-called homogeneous Malmquist bias [110]. The so-called non-homogeneous Malmquist bias [111], which appears when the density of objects is not constant, is usually overlooked. This approximation, certainly valid for the current data and given the estimates on the uncertainty of H_0 , should be tested for future analyses. Is worth noticing, as evident from the True Flagship results, that this effect is mitigated until it almost disappears by the presence of galaxy

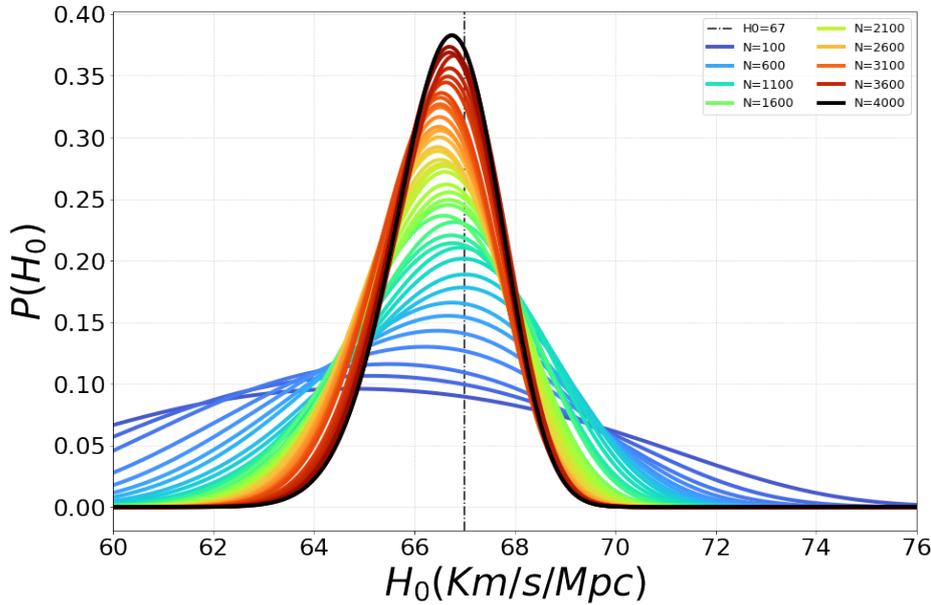


Figure 5.19. The full run with $N = 4000$ DSs. All events have $\sigma_{DL} = 10\%$, and $\Omega_{90} = 10 \text{ deg}^2$. The events are such that the research box is entirely contained in the catalogue, for every value of H_0 tested and are cross-correlated with the Uniform-Flagship catalogue. The image shows the evolution of the posterior when increasing the number of events. The plot shows different stages of the posterior, from the less informative blue curves toward the more informative ones in red, with the black curve been the total posterior.

clustering, since the spatial correlation of galaxies strongly helps the convergence of this method. The inclusion of this type of effect is under development and requires weighting the hosts, not with an average weight given by $N(z)$ but by the local density of the galaxy sample.

5.7 Forecast: 5 years of ET observation cross-matched with the Euclid spectroscopic galaxy catalogue

The scope of this final section is to obtain more realistic forecasts for the measurement of H_0 by combining a future dataset of dark sirens generated by the next-generation ET interferometer with a wide spectroscopic galaxy survey, mimicking the one being carried out by the Euclid satellite mission.

To achieve this, I defined a reference mock DS sample to be used as a baseline. The outcome of a GW observational campaign is a set of DSs detected above a given signal-to-noise ratio, expected to contain objects with different uncertainties in distance and angular positions. For the likelihood analyses it is often convenient to consider only a subset of DSs since low signal-to-noise events may add noise or bias to the results.

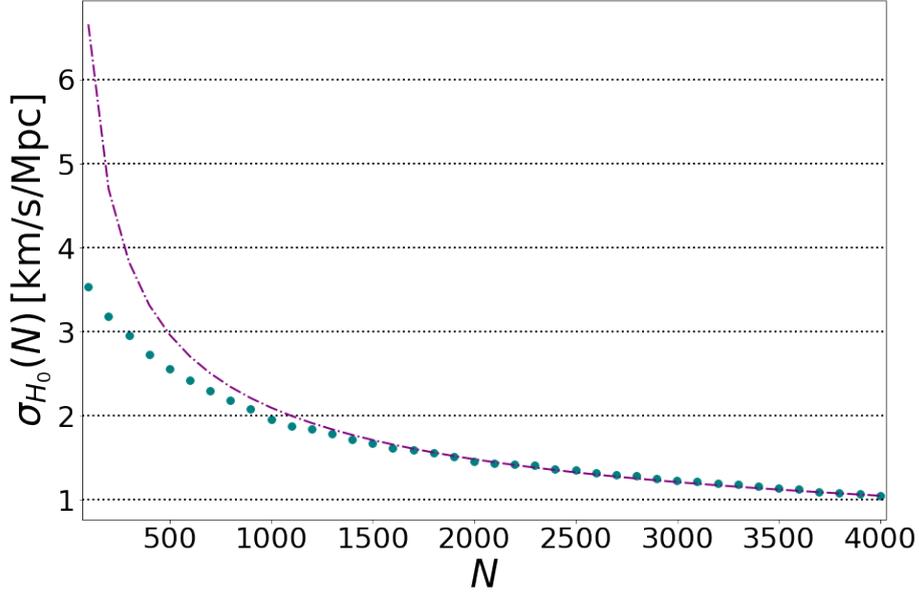


Figure 5.20. This figure shows the standard deviation σ_{H_0} as a function of the number of DS events. The dotted purple line is a curve proportional to $67/\sqrt{N}$.

In this analysis, the optimisation involves finding the best compromise between the number of events (N) and the precision of the measured DS quantities (σ_{DL} and Ω_{90}). Since the results are robust to variations in Ω_{90} I fix its value to 10 deg^2 and play with the remaining parameters.

The results are summarised in Figure 5.22 where I show that two combinations, $N = 300$, $\sigma_{DL} = 5\%$ and $N = 1000$, $\sigma_{DL} = 10\%$ give similar results, indicating that restricting the analysis to 30 % of the DS with a luminosity distance measured with higher precision gives comparable results to using a sample thrice as large with an error on the distance twice big. Therefore, I choose the former as the baseline, saving computational time over the latter.

Let's focus on the forecast for H_0 . According to [81], the Einstein Telescope is expected to observe between 100 and 400 GW events each year with a luminosity distance uncertainty of 5 % in the flagship redshift range. This is compatible with the choice for the baseline properties of the DSs catalogue exposed above. For the analysis, I will consider the optimistic scenario and use a reference sample of 2000 DS collected over 5 years of observations. I will assume that these DSs are located in an octant of the sky and have observed luminosity distances in the range of 7000 to 12000 Mpc, which overlaps with the redshift range spanned by the Euclid spectroscopic survey ($z = [0.9, 1.9]$). Additionally, more pessimistic scenarios with a lower number of detected GW events will also be considered.

I will first take a conservative approach and consider the case of an unclustered sample of hosts and DSs, hence ignoring the spatial correlation properties of the two

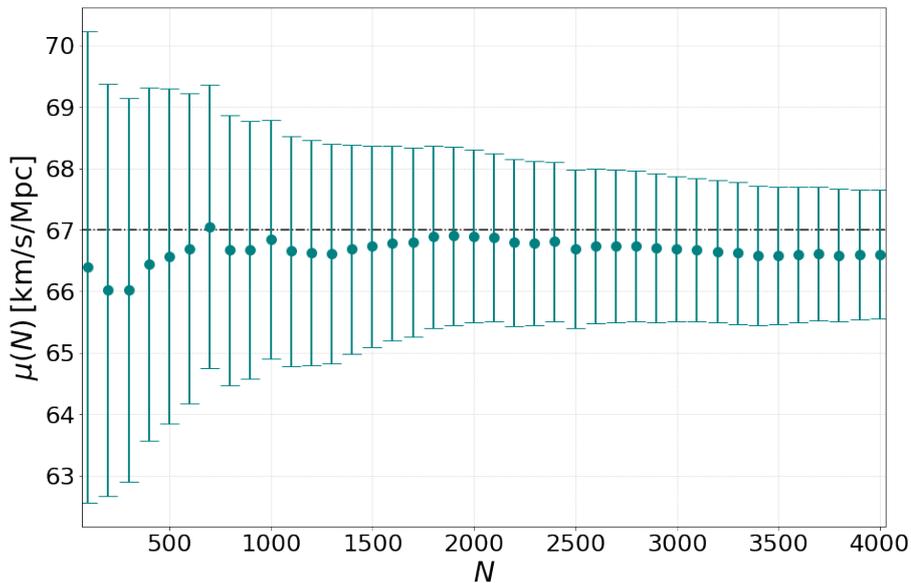


Figure 5.21. The evolution of the mean μ Vs N . The DSs are drawn by the Uniform-ideal catalogue and cross correlated with Uniform-flag.

populations. For the hosts I will consider the Uniform-Flagship catalogue which matches the expected $N(z)$ of the Euclid galaxies and is diluted to match their number density. First, a catalogue Uniform in comoving volume is generated, i.e. the Uniform-ideal catalogue. A subsample is then extracted from this catalogue, matching the $N(z)$ of the Flagship, making sure that the most populated bin in the Flagship and the extracted catalogue have the same number of hosts. Then the extracted catalogue is diluted to match the real Euclid density. Although the hosts will follow an overall distribution that is $N(z)$, there is no spatial correlation between the hosts, therefore we refer to this case as homogeneous. The DS catalogue will be taken from the parent Uniform-ideal catalogue, where not all DS are expected to have a matching host in the analysis.

The resulting posterior distribution for H_0 is depicted in Fig. 5.23. Remarkably, the expected uncertainty is very small, on the order of 1 %, even without spatial correlation between hosts. The sheer amount of DSs and their properties results in a good estimation of H_0 but, as will be evident later, the clustering drastically reduce the amount of DSs required to achieve this precision. Moreover, the estimate is unbiased, differing from the expected value (represented by the vertical dashed line) by less than the standard deviation of the quasi-Gaussian curve. Since the error is dominated by Poisson noise, we anticipate that even smaller DS samples, which may become available during the observational campaign, will be capable of providing stringent constraints on H_0 before the survey's completion.

This is further illustrated in Fig. 5.24, where I present the mean of the posterior

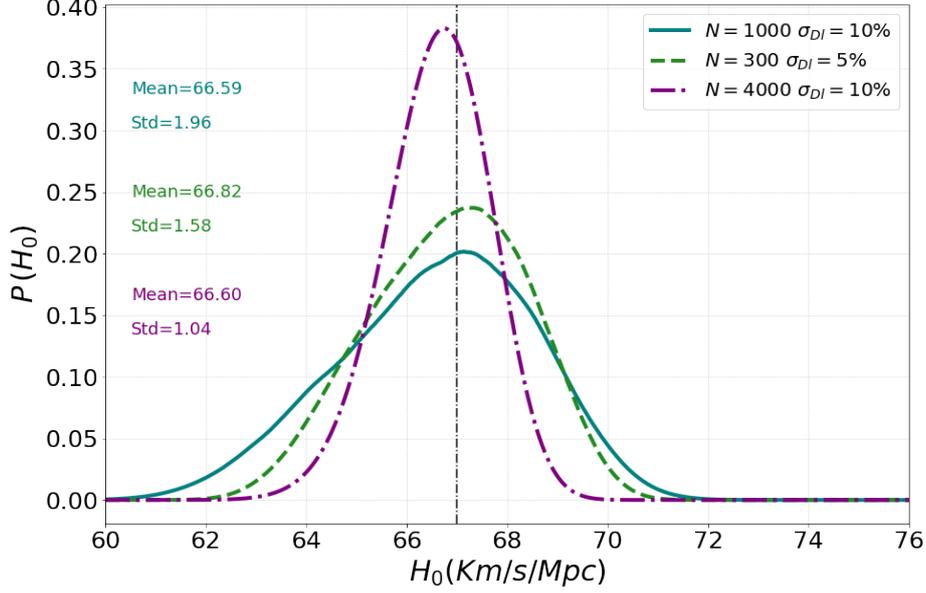


Figure 5.22. Comparison between the full run with $N = 4000$ DSs with $\sigma_{DL} = 10\%$, and $\Omega_{90} = 10 \text{ deg}^2$ in purple, a subset of this same run in teal and a run with 300 DSs with $\sigma_{DL} = 5\%$, and $\Omega_{90} = 10 \text{ deg}^2$. The events are such that the research box is entirely contained in the catalogue, for every value of H_0 tested and are cross-correlated with the Uniform-Flagship catalogue.

distribution of H_0 (dots) along with its standard deviation (error bars) as a function of N . It's observed that a high precision on the measurement of H_0 , comparable to that of the full sample, appears achievable with 1000 events, i.e., with only half of the full sample. The fact that the mean of the curve is consistently below the expected value, indicated by the dashed horizontal line, is not statistically significant, as the mismatch is below 1σ , and the points are highly correlated. The amplitude of the mismatch decreases with N suggesting that is not a predominant effect. The fact that each run approaches the true value of the Hubble constant from below, suggest that, even if below 1σ there is an offset. This offset can be caused by two main reason: either the extraction of the DSs is not totally under control or the normalisation function β do not compensate exactly for volume-like effects. The evolution of the standard deviation is shown in Fig. 5.25.

In the final step, I accounted for the clustering of both DSs and galaxies and their cross-correlation. To achieve this, I utilised the Flagship catalogue of mock galaxies randomly diluted to match the expected Euclid's galaxy density. This diluted catalogue served as the host galaxy catalogue used in the analysis. Similarly, the catalogue of DS was randomly selected from the Flagship catalogue. This choice corresponds to an optimistic scenario in which a significant fraction of DS have their host galaxy in the matching galaxy sample. To avoid introducing this bias, I used the procedure described in Section 5.2 and displaced the position of each DS

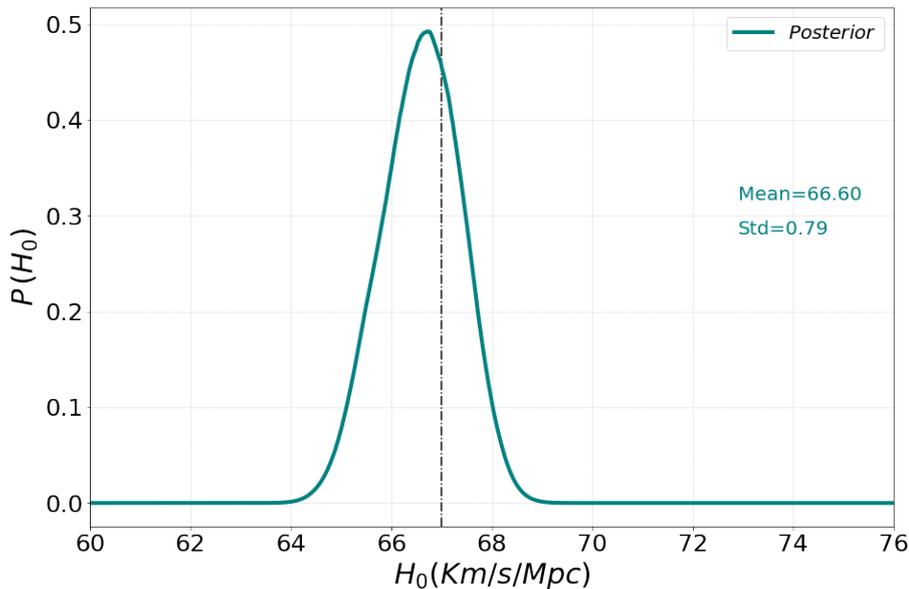


Figure 5.23. H_0 posterior distribution for a set of 2000 DSs from the Uniform-ideal catalogue and cross correlated with the Uniform-Flagship. This will be the new baseline for the rest of the section. All DSs have $\sigma_{DL} = 5\%$, and $\Omega_{90} = 10 \text{ deg}^2$.

around the host flagship galaxy according to the galaxy-galaxy correlation function in eq. (5.1), to account for the probability of a galaxy being at a separation r from another one in the sample.

The resulting posterior distribution is displayed in Fig. 5.26 (continuous curves) and compared to the one of the uniform case (dotted curve) previously shown in Fig. 5.23. In particular, the posterior in purple in Fig. 5.26 appears to be perfectly centred, not approaching the true value from below as has been the case until now. There is currently no precise explanation for the absence of any bias, as the criterion to determine if a run is biased has always been to check that the true value is within one standard deviation. For now, the most likely hypothesis is that it was a particularly fortunate run, as also indicated by the curves in Fig. 5.29. However, this suggests that the non-homogeneous Malmquist bias is negligible even at these sensitivities.

Regarding the width of the posterior, the result is very encouraging as it demonstrates that, as mentioned earlier, it is not necessary to have the host in the sample because DSs and their hosts are spatially correlated with galaxies. Moreover, the greater the correlation between not only the spatial but also the physical properties of the sources and the galaxies, the better the estimate of H_0 will be. For example, if it were discovered that brighter DSs correlate better with massive galaxies, then the determination of the Hubble constant could be further improved by studying only a subsample of intrinsically brighter DSs.

It is evident that accounting for the cross-correlation between DSs and galaxies

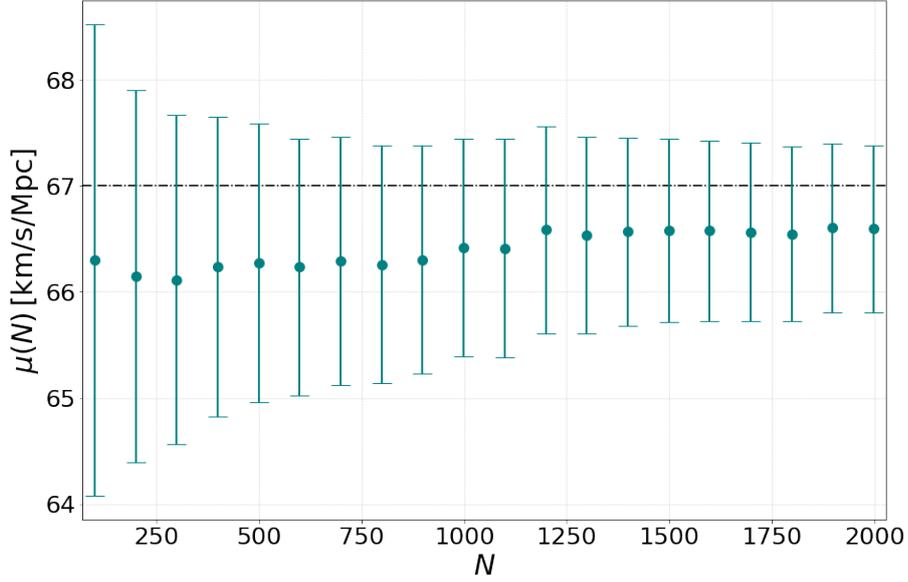


Figure 5.24. In teal, the evolution of the mean μ of the posterior for the increasing number of events analysed with the true flagship catalogue.

reduces the uncertainty on H_0 by a factor of 2. This improvement is genuine and induced by clustering, as it does not depend on the number of DS, as shown in Fig. 5.26. The factor of 2 reduction in the measured variance remains roughly constant across the entire N -range, as clear from Fig. 5.27.

Accounting for clustering does not introduce any bias since the mean of the posterior distribution is consistent with the expected H_0 value as illustrated in Fig. 5.28 which shows the mean values (dots) and variances (error bars) of the posteriors obtained for DS samples of increasing sizes and compare the results with those obtained with the unclustered sample (offset asterisks)

Taking Fig. 5.27 and Fig 5.28 as references, it can be verified that, upon reaching about 1000 events, this strategy becomes competitive with the statistical uncertainties in measuring H_0 obtained from local distance indicators and the CMB, that have uncertainty of $\sim 1\%$ [16, 17]. Given the estimated rate of gold events, ranging from 100 to 400, where 100 is a very pessimistic estimate, it is plausible that between 2 and 5 years of ET+Euclid data are necessary to be competitive, or in the optimistic case, better than the available estimates. This technique, in both the optimistic and non-optimistic scenario, will enhance the statistical significance of the Hubble tension. If for example the obtained value for H_0 (H_0^{GW}) with this technique equals the one from the local Universe, the statistical significance of the tension will be no

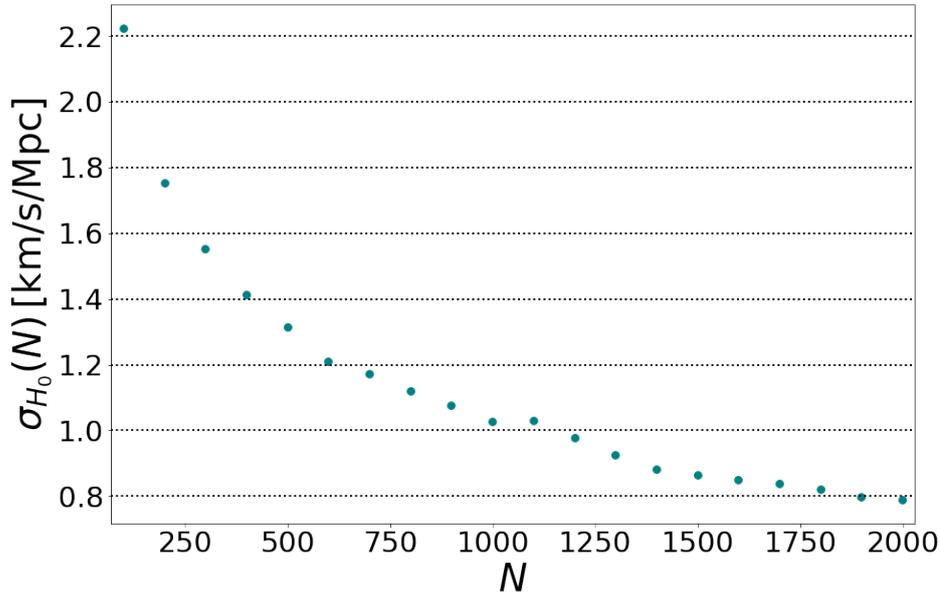


Figure 5.25. In teal, the evolution of the standard deviation σ_{H_0} of the posterior for the increasing number of events.

less than 7σ given that in the pessimistic case

$$\frac{H_0^{GW} - H_0^{CMB}}{\sqrt{\sigma_{GW}^2 + \sigma_{CMB}^2}} = \frac{5.63}{\sqrt{0.50^2 + 0.62^2}} = 7.07 \quad (5.6)$$

If H_0^{GW} coincides with the CMB one, than in the pessimistic case the statistical significance of the tension is $\sim 5.5\sigma$.

It is worth stressing that this result is robust to the choice of the selected sample. To check this, I have repeated the same analysis using two different samples, independently selected from the Flagship catalogue.

All the posterior distributions, shown in Fig. 5.29, look similar and consistent with a sub-percent precision in the measurement of H_0 . This preliminary test reinforce the belief that the strategy is well posed and the simulations are correctly reproducing the physics of the problem. The presence of a smaller offset in this kind of runs with a non-homogeneous host catalogue suggest that the cause of the bias is in the selection of the DSs and further investigations on this aspect will be done.

In conclusion, given a number of gold events between one thousand and two thousand, spread over five years of data taking, by correlating a set of ET-like DSs and the Euclid flagship it is possible to determine H_0 with an accuracy of between 1% and 0.5%. This result was obtained under certain assumptions: detector stationary during Earth orbit, no correlation between the quantities involved etc. However, the analysis showed that a fundamental actor is the clustering of galaxies and the

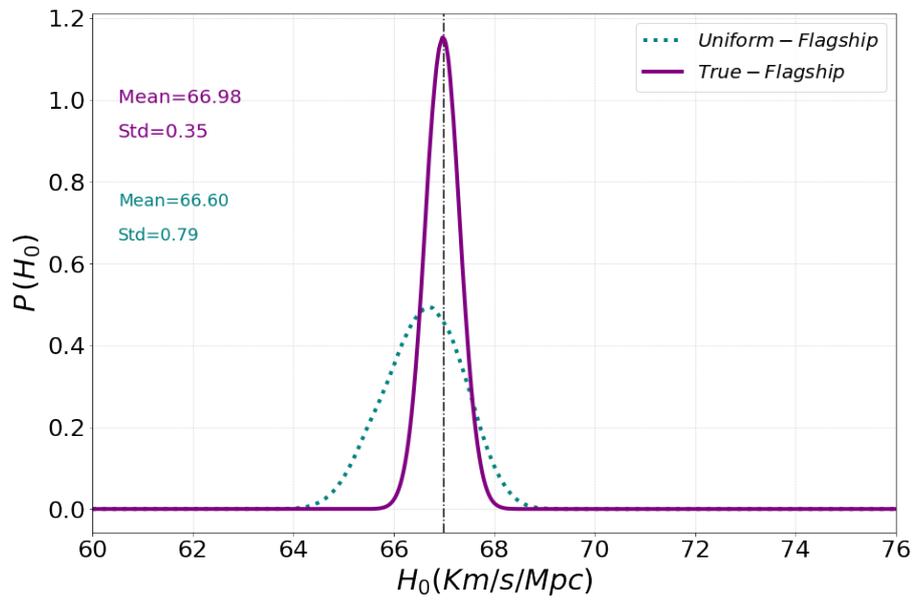


Figure 5.26. In teal, the posterior from a collection of 2000 DSs sampled from the real flagship and cross correlated with the Euclid mock data. In purple, the posterior 2000 DSs extracted from a uniform distribution and cross correlated with the true flagship. The DSs are extracted between an observed $DL_{min} = 7000$ Mpc and a $DL_{min} = 12000$ Mpc.

possibility of determining the redshift of the hosts spectroscopically. This makes it possible to achieve H_0 estimates competitive with the major experiments already mentioned in the text.

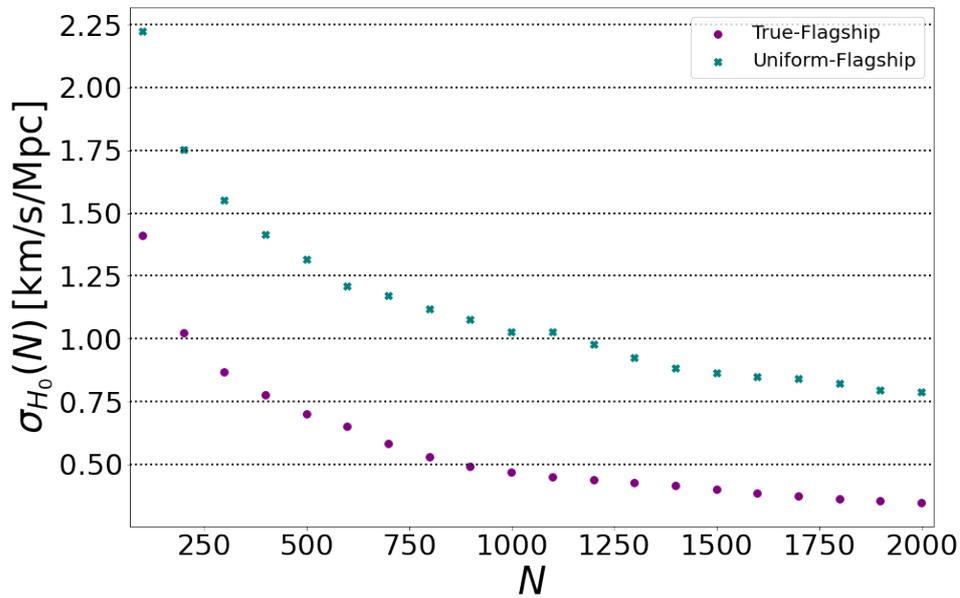


Figure 5.27. In teal, the evolution of the standard deviation σ_{H_0} of the posterior for the increasing number of events analysed with the true flagship catalogue for the aforementioned Dl range. In purple, the same quantity but analysed with a uniform catalogue with the same $N(z)$ as the true flagship. The final posteriors are shown in Fig. 5.26

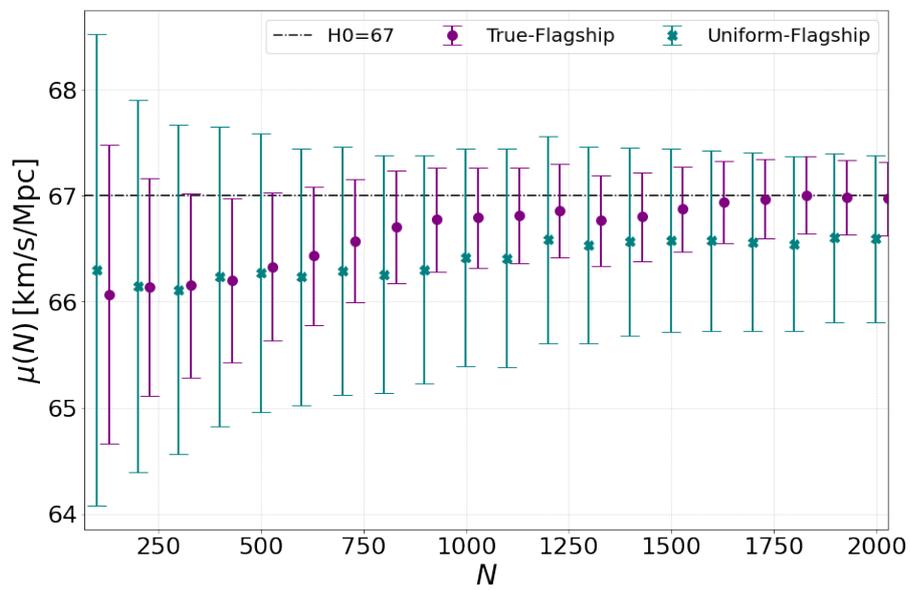


Figure 5.28. In teal, the evolution of the mean μ of the posterior in Fig. 5.26 for the increasing number of events cross-correlated with the true flagship catalogue. In purple, the same quantity but the DSs are drawn from a uniform distribution and cross correlated with the Uniform-Flagship catalogue.

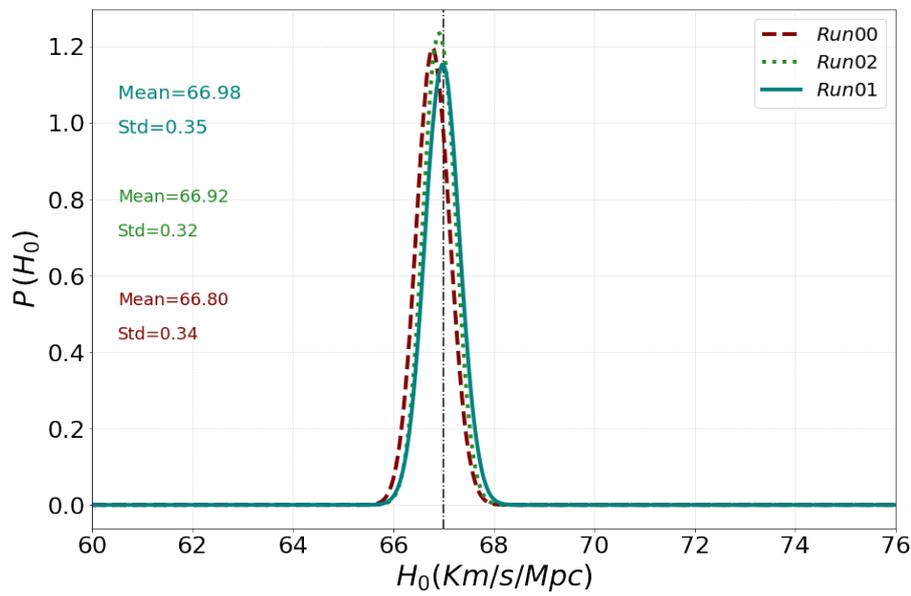


Figure 5.29. In teal the posterior from a collection of 2000 DSs sampled from the real flagship and cross correlated with the Euclid mock data, the true flagship. In green, the posterior given by another set of DSs named Run02 and in maroon, the third posterior for the third set of DSs. The DSs are extracted between an observed $Dl_{min} = 7000$ Mpc and a $Dl_{min} = 12000$ Mpc. All DSs have $\sigma_{DL} = 5\%$, and $\Omega_{90} = 10 \text{ deg}^2$.

Conclusions

The detection of gravitational waves has indeed marked a significant milestone in physics and astronomy. Not only do they provide further validation of Einstein's theory of gravity, but they also introduce a new means of exploring the cosmos beyond traditional electromagnetic radiation.

Whether GWs will revolutionise cosmology depends on their potential to serve as high-precision cosmological probes, addressing key questions such as the Hubble tension – the discrepancy between measurements of H_0 at the last scattering epoch and those from local measurements using distance indicators.

GWs offer a promising avenue for independently probing H_0 and its evolution by targeting sources at low to moderate redshifts using current and future detectors. This is best achieved through the observation of "bright sirens," where both the distance of the source from the GW signal and the redshift of the galaxy host can be measured accurately from the GW signal and the observed photon flux. Indeed, the scarcity of bright sirens observed so far presents a significant challenge to using them as the primary means of measuring H_0 . As a result, an alternative hybrid strategy is being explored, which involves combining dark GW events with the known positions of bright galaxies. Success with this approach hinges on statistical considerations: a substantial number of dark sirens must be detected within the same volume where the spatial positions of a large number of potential galaxy hosts have been accurately mapped. While this approach is not currently feasible with the available datasets, the future looks promising. The number of dark GW events detected by existing interferometers is gradually increasing and is expected to skyrocket with the advent of next-generation detectors like ET and CE. Additionally, ongoing large observational campaigns such as DESI and Euclid are continuously mapping the distribution of galaxies across ever-expanding portions of the observable Universe. These developments provide hope that the hybrid strategy will become increasingly viable in the future.

This thesis focuses on exploring the potential of using dark sirens as alternative probes for H_0 with comparable precision to existing methods. To achieve this goal, I developed a prototype end-to-end pipeline called CODE (Cross cOrrelation with Dark sirEns). The CODE pipeline is designed to generate mock dark sirens and galaxy catalogues that emulate the characteristics of future datasets, such as those expected from next-generation GW detectors like ET, and large observational campaigns to map galaxies like Euclid. Using these simulated datasets, the pipeline employs hierarchical Bayesian inference approach to estimate H_0 .

The flexibility of the CODE architecture allows it to handle various types of datasets, including existing ones. While the current implementation focuses on

exploring the combination of simulated ET-like and Euclid-like catalogues of dark sirens and galaxies, the pipeline can be adapted to accommodate other datasets as well.

The validation process involved running several tests using ideal datasets to ensure the robustness of the pipeline. These tests aimed to assess the impact of various factors on the analysis outcome, including uncertainties in the luminosity distance and angular position of the dark sirens. A notable aspect of this validation was its focus on dark sirens at moderate redshifts, within the range covered by the Euclid spectroscopic survey, $0.8 < z < 2$, which is a departure from previous studies that primarily examined local samples due to the limited sensitivity of current gravitational wave detectors.

An intriguing and somewhat surprising finding from the analysis is the limited impact of uncertainties in the angular position of the dark sirens on the estimation of H_0 . This observation could be linked to the dense distribution of potential hosts in the galaxy catalogue. It's important to emphasise that this result was derived using catalogues of unclustered and entirely independent objects, both galaxies and dark sirens. Given its potential significance in informing future survey strategies, it warrants further investigation using more realistic datasets.

The precision of the dark sirens' distance estimates plays a crucial role in the analysis, highlighting the potential for obtaining competitive estimates of H_0 by focusing on a "golden sample" of dark sirens with the most accurately measured distances. This underscores the importance of accurately measuring the redshifts of the host galaxies, preferably with matching or superior precision. Consequently, this type of analysis necessitates the use of galaxy catalogues with redshifts determined from spectroscopic measurements, as photometric redshifts derived from flux measurements across different bands typically have errors comparable to those of the dark sirens' distances.

The testing campaign also provided reassuring confirmation that errors on H_0 are adequately described by a simple Poisson model when the number of dark sirens in the sample is large. This insight is very useful for determining the survey strategy and duration, providing guidance for optimising observational efforts. The application of the pipeline to more realistic datasets involved ensuring that the radial distribution and galaxy density for the host galaxies matched real observations, and accurately representing the number, radial distribution, and angular and distance errors of the dark DS. The results of the analysis suggest that a 5-year observational campaign with the next-generation ET interferometer, combined with cross-matching with the Euclid spectroscopic galaxy survey expected to be completed by 2029, can achieve a percent precision and accuracy in the measurement of H_0 .

Accounting for the clustering properties of both the DS and the galaxies in the analysis further improves the precision of the H_0 estimate, without introducing significant systematic errors. Indeed, this prediction represents an optimistic outlook, as it overlooks many potential sources of error that could affect the analysis at the sub-percent level. Nonetheless, it unequivocally highlights the potential of dark sirens as high-precision cosmological probes. These promising results encourage further analysis and exploration along three main directions.

First of all, I plan to enhance the realism of the analysis, I aim to thoroughly incorporate all potential sources of random and systematic errors at various stages

of the pipeline. For instance, the angular position error, which I previously modelled as a simple 2D Gaussian function on the celestial sphere, needs refinement. Existing DS events often exhibit elongated and irregular shapes in their 2D probability areas, which are better represented using healpix probability maps [112]. While the CODE pipeline is already equipped to generate DS events complemented by healpix probability maps, further development is required to propagate this information to the likelihood analysis.

Another area for improvement lies in enhancing the "observation" steps, which connect the intrinsic properties of the gravitational wave source to the observed DS event as captured by the interferometer. By refining this process, I can achieve a more self-consistent distribution for the observed DS properties, including their distance errors. Furthermore, it's essential to accurately simulate the geometry of the galaxy survey, accounting for features like holes and masked areas, to properly address edge effects.

The second improvement I aim to focus on for CODE is its adaptability to handle diverse datasets. As mentioned earlier, the architecture of the pipeline is sufficiently flexible to simulate and conduct Bayesian inference analyses on various sets of dark sirens and galaxies. Specifically, I am keen on leveraging CODE to scrutinize the existing LISA-VIRGO DS events and those from forthcoming campaigns conducted by the upgraded LIGO-Virgo-Kagra interferometer system. On the galaxy front, I intend to incorporate the emission line galaxies' catalogue being generated by the DESI survey.

Last, but not least, I intend to expand beyond the measurement of H_0 . Finke et al. in [6] have already extended the hierarchical Bayesian framework to accommodate models of modified gravity. I plan to replicate their approach and further extend it to enable the constraint of additional cosmological parameters using the dark sirens datasets.

Bibliography

- [1] J. Amiaux, R. Scaramella, Y. Mellier, B. Altieri, C. Burigana, Antonio Da Silva, P. Gomez, J. Hoar, R. Laureijs, E. Maiorano, D. Magalhães Oliveira, F. Renk, G. Saavedra Criado, I. Tereno, J. L. Auguères, J. Brinchmann, M. Cropper, L. Duvet, A. Ealet, P. Franzetti, B. Garilli, P. Gondoin, L. Guzzo, H. Hoekstra, R. Holmes, K. Jahnke, T. Kitching, M. Meneghetti, W. Percival, and S. Warren. Euclidmission: building of a reference survey. In Mark C. Clampin, Giovanni G. Fazio, Howard A. MacEwen, and Jacobus M. Oschmann, editors, *SPIE Proceedings*. SPIE, August 2012.
- [2] Michael E et al. The dark energy spectroscopic instrument (desi), 2019.
- [3] Robert T. Zellem, Bijan Nemati, Guillermo Gonzalez, Marie Ygouf, Vanessa P. Bailey, Eric J. Cady, M. Mark Colavita, Sergi R. Hildebrandt, Erin R. Maier, Bertrand Mennesson, Lindsey Payne, Neil Zimmerman, Ruslan Belikov, Robert J. De Rosa, John Debes, Ewan S. Douglas, Julien Girard, Tyler Groff, Jeremy Kasdin, Patrick J. Lowrance, Bruce Macintosh, Daniel Ryan, and Carey Weisberg. Nancy grace roman space telescope coronagraph instrument observation calibration plan, 2022.
- [4] B. P. Abbott et al. Gravitational waves and gamma-rays from a binary neutron star merger: Gw170817 and grb 170817a. *The Astrophysical Journal Letters*, 848(2):L13, oct 2017.
- [5] Edwin Hubble. A relation between distance and radial velocity among extragalactic nebulae. *Proceedings of the National Academy of Sciences*, 15(3):168–173, 1929.
- [6] Andreas Finke, Stefano Foffa, Francesco Iacovelli, Michele Maggiore, and Michele Mancarella. Cosmology with LIGO/virgo dark sirens: Hubble parameter and modified gravitational wave propagation.
- [7] W. Ballard, A. Palmese, I. Magaña Hernandez, S. BenZvi, J. Moon, A. J. Ross, G. Rossi, J. Aguilar, S. Ahlen, R. Blum, D. Brooks, T. Claybaugh, A. de la Macorra, A. Dey, P. Doel, J. E. Forero-Romero, S. Gontcho A Gontcho, K. Honscheid, A. Kremin, M. Manera, A. Meisner, R. Miquel, J. Moustakas, F. Prada, E. Sanchez, G. Tarlé, and Z. Zhou. A dark siren measurement of the hubble constant with the ligo/virgo gravitational wave event gw190412 and desi galaxies, 2023.

- [8] V. Alfradique, C. R. Bom, A. Palmese, G. Teixeira, L. Santana-Silva, A. Drlica-Wagner, A. H. Riley, C. E. Martínez-Vázquez, D. J. Sand, G. S. Stringfellow, G. E. Medina, J. A. Carballo-Bello, Y. Choi, J. Esteves, G. Limberg, B. Mutlu-Pakdil, N. E. D. Noël, A. B. Pace, J. D. Sakowska, and J. F. Wu. A dark siren measurement of the hubble constant using gravitational wave events from the first three ligo/virgo observing runs and delve, 2023.
- [9] Nicola Borghi, Michele Mancarella, Michele Moresco, Matteo Tagliazucchi, Francesco Iacovelli, Andrea Cimatti, and Michele Maggiore. Cosmology and astrophysics with standard sirens and galaxy catalogs in view of future gravitational wave observations. *The Astrophysical Journal*, 964(2):191, March 2024.
- [10] Albert Einstein. Cosmological Considerations in the General Theory of Relativity. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)*, 1917:142–152, 1917.
- [11] Bradley W. Carroll and Dale A. Ostlie. *An introduction to modern astrophysics and cosmology*. 2006.
- [12] Georges Lemaître. L’Univers en expansion. *Annales de la Société Scientifique de Bruxelles*, 53:51, January 1933.
- [13] Eleonora Di Valentino et al. In the realm of the hubble tension.
- [14] Richard I. Anderson, Nolan W. Koblischke, and Laurent Eyer. Reconciling astronomical distance scales with variable red giant stars.
- [15] R. Brent Tully. The hubble constant: A historical review, 2023.
- [16] Yuhei S. Murakami, Adam G. Riess, Benjamin E. Stahl, W. D’Arcy Kenworthy, Dahne-More A. Pluck, Antonella Macoretta, Dillon Brout, David O. Jones, Dan M. Scolnic, and Alexei V. Filippenko. Leveraging sn ia spectroscopic similarity to improve the measurement of h_0 . *Journal of Cosmology and Astroparticle Physics*, 2023(11):046, nov 2023.
- [17] Planck Collaboration. Planck 2018 results - vi. cosmological parameters. *A&A*, 641:A6, 2020.
- [18] B.F. Schutz. Determining the Hubble constant from gravitational wave observations. *nat*, 323(6086):310–311, September 1986.
- [19] B.P et all Abbott. Observation of gravitational waves from a binary black hole merger. *Physical Review Letters*, 116(6), February 2016.
- [20] Albert Einstein. Näherungsweise Integration der Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, pages 688–696, January 1916.
- [21] B. P. Abbott et al. Multi-messenger observations of a binary neutron star merger*. *The Astrophysical Journal Letters*, 848(2):L12, oct 2017.

- [22] Abbott et al. Gwtc-2: Compact binary coalescences observed by ligo and virgo during the first half of the third observing run. *Physical Review X*, 11(2), Jun 2021.
- [23] Alam et al. Completed sdss-iv extended baryon oscillation spectroscopic survey: Cosmological implications from two decades of spectroscopic surveys at the apache point observatory. *Physical Review D*, 103(8), April 2021.
- [24] Dawson et al. The baryon oscillation spectroscopic survey of sdss-iii. *The Astronomical Journal*, 145(1):10, December 2012.
- [25] et al Dawson. The sdss-iv extended baryon oscillation spectroscopic survey: Overview and early data. *The Astronomical Journal*, 151(2):44, February 2016.
- [26] Daniel J. Eisenstein, Hee-Jong Seo, and Martin White. On the robustness of the acoustic scale in the low-redshift clustering of matter. *The Astrophysical Journal*, 664:660–674, August 1, 2007 2007. n/a.
- [27] Scaramella et al. Euclid preparation: I. the euclid wide survey. *Astronomy & Astrophysics*, 662:A112, June 2022.
- [28] Misao Sasaki, Teruaki Suyama, Takahiro Tanaka, and Shuichiro Yokoyama. Primordial black holes—perspectives in gravitational wave astronomy. *Classical and Quantum Gravity*, 35(6):063001, March 2018.
- [29] Abbott et al. Gw170817: Measurements of neutron star radii and equation of state. *Physical Review Letters*, 121(16), October 2018.
- [30] Emanuele Berti, Vitor Cardoso, and Paolo Pani. Breit-wigner resonances and the quasinormal modes of anti-de sitter black holes. *Physical Review D*, 79(10), May 2009.
- [31] Vitor Cardoso and Arianna Foschi. Geodesic structure and quasinormal modes of a tidally perturbed spacetime. *Physical Review D*, 104(2), July 2021.
- [32] Caio F.B. Macedo, Vitor Cardoso, Luís C.B. Crispino, and Paolo Pani. Quasinormal modes of relativistic stars and interacting fields. *Physical Review D*, 93(6), March 2016.
- [33] Michele Maggiore, Chris Van Den Broeck, Nicola Bartolo, Enis Belgacem, Daniele Bertacca, Marie Anne Bizouard, Marica Branchesi, Sebastien Clesse, Stefano Foffa, Juan García-Bellido, Stefan Grimm, Jan Harms, Tanja Hinderer, Sabino Matarrese, Cristiano Palomba, Marco Peloso, Angelo Ricciardone, and Mairi Sakellariadou. Science case for the einstein telescope. *Journal of Cosmology and Astroparticle Physics*, 2020(03):050–050, March 2020.
- [34] Henrietta S. Leavitt and Edward C. Pickering. Periods of 25 Variable Stars in the Small Magellanic Cloud. *Harvard College Observatory Circular*, 173:1–3, March 1912.
- [35] Henrietta S. Leavitt. 1777 variables in the Magellanic Clouds. *Annals of Harvard College Observatory*, 60:87–108.3, January 1908.

- [36] Brian P. Schmidt, Nicholas B. Suntzeff, M. M. Phillips, Robert A. Schommer, Alejandro Clocchiatti, Robert P. Kirshner, Peter Garnavich, Peter Challis, B. Leibundgut, J. Spyromilio, Adam G. Riess, Alexei V. Filippenko, Mario Hamuy, R. Chris Smith, Craig Hogan, Christopher Stubbs, Alan Diercks, David Reiss, Ron Gilliland, John Tonry, Jose Maza, A. Dressler, J. Walsh, and R. Ciardullo. The high- z supernova search: Measuring cosmic deceleration and global curvature of the universe using type ia supernovae. *The Astrophysical Journal*, 507(1):46–63, November 1998.
- [37] D. J. Fixsen. The temperature of the cosmic microwave background. *The Astrophysical Journal*, 707(2):916, nov 2009.
- [38] Gary Steigman. Primordial alchemy: From the big bang to the present universe, 2002.
- [39] M. Fukugita, C. J. Hogan, and P. J. E. Peebles. The cosmic baryon budget. *The Astrophysical Journal*, 503(2):518, aug 1998.
- [40] Masataka Fukugita and P. J. E. Peebles. The cosmic energy inventory. *The Astrophysical Journal*, 616(2):643, dec 2004.
- [41] F. Nicastro, J. Kaastra, Y. Krongold, S. Borgani, E. Branchini, R. Cen, M. Dadina, C. W. Danforth, M. Elvis, F. Fiore, A. Gupta, S. Mathur, D. Mayya, F. Paerels, L. Piro, D. Rosa-Gonzalez, J. Schaye, J. M. Shull, J. Torres-Zafra, N. Wijers, and L. Zappacosta. Observations of the missing baryons in the warm-hot intergalactic medium. *Nature*, 558(7710):406–409, June 2018.
- [42] Heinz Andernach and Fritz Zwicky. English and spanish translation of zwicky’s (1933) the redshift of extragalactic nebulae, 2017.
- [43] Vera C. Rubin and Jr. Ford, W. Kent. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *apj*, 159:379, February 1970.
- [44] J.A. Peacock. *Cosmological Physics*. Cambridge Astrophysics. Cambridge University Press, 1999.
- [45] Scott Dodelson. *Modern Cosmology*. Academic Press, Elsevier Science, 2003.
- [46] Marc Schumann. Direct detection of wimp dark matter: concepts and status. *Journal of Physics G: Nuclear and Particle Physics*, 46(10):103003, aug 2019.
- [47] Bernard Carr and Florian Kühnel. Primordial black holes as dark matter candidates. *SciPost Physics Lecture Notes*, May 2022.
- [48] Monique Signore and Denis Puy. Supernovae and cosmology. *New Astronomy Reviews*, 45(4–5):409–423, March 2001.
- [49] Alexei V. Filippenko and Adam G. Riess. Results from the high- z supernova search team. *Physics Reports*, 307(1–4):31–44, December 1998.
- [50] Joshua A. Frieman, Michael S. Turner, and Dragan Huterer. Dark energy and the accelerating universe. *Annual Review of Astronomy and Astrophysics*, 46(1):385–432, September 2008.

- [51] Florian Wolz. *Gravitational Closure of Matter Field Equations: General Theory and Perturbative Solutions*. PhD thesis, 05 2022.
- [52] Jenny T Wan, Adam B Mantz, Jack Sayers, Steven W Allen, R Glenn Morris, and Sunil R Golwala. Measuring h_0 using x-ray and sz effect observations of dynamically relaxed galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 504(1):1062–1076, April 2021.
- [53] Tommaso Treu, Sherry H. Suyu, and Philip J. Marshall. Strong lensing time-delay cosmography in the 2020s. *The Astronomy and Astrophysics Review*, 30(1), November 2022.
- [54] Adam G. Riess, Stefano Casertano, Wenlong Yuan, J. Bradley Bowers, Lucas Macri, Joel C. Zinn, and Dan Scolnic. Cosmic distances calibrated to 1% precision with gaiaedr3 parallaxes and hubble space telescope photometry of 75 milky way cepheids confirm tension with Λ cdm. *The Astrophysical Journal Letters*, 908(1):L6, feb 2021.
- [55] Licia Verde, Nils Schöneberg, and Héctor Gil-Marín. A tale of many h_0 , 2023.
- [56] Andrew R. Liddle and David H. Lyth. *Cosmological Inflation and Large-Scale Structure*. Cambridge University Press, 2000.
- [57] C. L. Bennett, A. Kogut, G. Hinshaw, A. J. Banday, E. L. Wright, K. M. Gorski, D. T. Wilkinson, R. Weiss, G. F. Smoot, S. S. Meyer, J. C. Mather, P. Lubin, K. Loewenstein, C. Lineweaver, P. Keegstra, E. Kaita, P. D. Jackson, and E. S. Cheng. Cosmic temperature fluctuations from two years of coBE differential microwave radiometers observations. *The Astrophysical Journal*, 436:423, December 1994.
- [58] Wayne Hu and Scott Dodelson. Cosmic microwave background anisotropies. *Annual Review of Astronomy and Astrophysics*, 40(1):171–216, September 2002.
- [59] Daniel J. Eisenstein and Wayne Hu. Baryonic features in the matter transfer function. *The Astrophysical Journal*, 496(2):605, apr 1998.
- [60] Daniel J. Eisenstein et al. Detection of the baryon acoustic peak in the large-scale correlation function of sdss luminous red galaxies. *The Astrophysical Journal*, 633(2):560, nov 2005.
- [61] et al Aubourg. Cosmological implications of baryon acoustic oscillation measurements. *Phys. Rev. D*, 92:123516, Dec 2015.
- [62] R. Brent Tully, Cullan Howlett, and Daniel Pomarède. Ho’oleilana: An individual baryon acoustic oscillation? *The Astrophysical Journal*, 954(2):169, sep 2023.
- [63] B. Trahin, L. Breuval, P. Kervella, A. Mérand, N. Nardetto, A. Gallenne, V. Hocdé, and W. Gieren. Inspecting the cepheid parallax of pulsation using gaiaedr3 parallaxes: Projection factor and period-luminosity and period-radius relations. *Astronomy& Astrophysics*, 656:A102, December 2021.

- [64] Adam G. et al Riess. A comprehensive measurement of the local value of the hubble constant with 1 km s⁻¹ mpc⁻¹ uncertainty from the hubble space telescope and the sh0es team. *The Astrophysical Journal Letters*, 934(1):L7, July 2022.
- [65] Wendy L. Freedman. Measurements of the hubble constant: Tensions in perspective*. *The Astrophysical Journal*, 919(1):16, September 2021.
- [66] Peter Garnavich, Charlotte M. Wood, Peter Milne, Joseph B. Jensen, John P. Blakeslee, Peter J. Brown, Daniel Scolnic, Benjamin Rose, and Dillon Brout. Connecting infrared surface brightness fluctuation distances to type ia supernova hosts: Testing the top rung of the distance ladder. *The Astrophysical Journal*, 953(1):35, aug 2023.
- [67] S. Dhawan, D. Brout, D. Scolnic, A. Goobar, A. G. Riess, and V. Miranda. Cosmological model insensitivity of local h0 from the cepheid distance ladder. *The Astrophysical Journal*, 894(1):54, may 2020.
- [68] R. A. Hulse and J. H. Taylor. Discovery of a pulsar in a binary system. *apjl*, 195:L51–L53, January 1975.
- [69] B.P. et al. Abbott. Observation of gravitational waves from a binary black hole merger. *Physical Review Letters*, 116(6), February 2016.
- [70] Pau Amaro-Seoane et al. Laser interferometer space antenna, 2017.
- [71] LIGO Scientific Collaboration. Advanced LIGO. *Classical and Quantum Gravity*, 32(7):074001, April 2015.
- [72] Virgo Collaboration. Advanced virgo: a second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, 32(2):024001, December 2014.
- [73] GEO Collaboration. Geo 600 and the geo-hf upgrade program: successes and challenges. *Classical and Quantum Gravity*, 33(7):075009, March 2016.
- [74] KAGRA Collaboration. Overview of kagra: Detector design and construction history, 2020.
- [75] I. Kondrashov, Dmitry Simakov, Farit Khalili, and Stefan Danilishin. Optimizing the regimes of advanced ligo gravitational wave detector for multiple source types. *Physical Review D*, 78, 07 2008.
- [76] Mauro Pieroni, Angelo Ricciardone, and Enrico Barausse. Detectability and parameter estimation of stellar origin black hole binaries with next generation gravitational wave detectors. *Scientific Reports*, 12, 10 2022.
- [77] S. Hild, Matthew Abernathy, Fausto Acernese, P. Amaro-Seoane, Nils Andersson, K. Arun, Fabrizio Barone, B. Barr, M. Barsuglia, Mark Beker, Nicola Beveridge, S. Birindelli, Samrat Bose, L. Bosi, S. Braccini, C Bradaschia, Tomasz Bulik, Enrico Calloni, Giancarlo Cella, and K Yamamoto. Sensitivity

- studies for third-generation gravitational wave observatories. *Classical and Quantum Gravity*, 28, 12 2010.
- [78] Quirijn Meijer, Melissa Lopez, Daichi Tsuna, and Sarah Caudill. Gravitational-wave searches for cosmic string cusps in einstein telescope data using deep learning. *Physical Review D*, 109(2), January 2024.
- [79] Bernard F Schutz. Networks of gravitational wave detectors and three figures of merit. *Classical and Quantum Gravity*, 28(12):125023, may 2011.
- [80] Piotr Jaranowski, Andrzej Królak, and Bernard F. Schutz. Data analysis of gravitational-wave signals from spinning neutron stars: The signal and its detection. *Phys. Rev. D*, 58:063001, Aug 1998.
- [81] Branchesi et al. Science with the einstein telescope: a comparison of different designs. *Journal of Cosmology and Astroparticle Physics*, 2023(07):068, July 2023.
- [82] Douglas Potter, Joachim Stadel, and Romain Teyssier. PKDGRAV3: beyond trillion particle cosmological simulations for the next era of galaxy surveys. *Computational Astrophysics and Cosmology*, 4(1), 2017.
- [83] Peter S. Behroozi, Risa H. Wechsler, and Hao-Yi Wu. The rockstar phase-space temporal halo finder and the velocity offsets of cluster cores. *The Astrophysical Journal*, 762(2):109, December 2012.
- [84] Jeremy Tinker, Andrey V. Kravtsov, Anatoly Klypin, Kevork Abazajian, Michael Warren, Gustavo Yepes, Stefan Gottlöber, and Daniel E. Holz. Toward a halo mass function for precision cosmology: The limits of universality. *The Astrophysical Journal*, 688(2):709–728, December 2008.
- [85] A COORAY and R SHETH. Halo models of large scale structure. *Physics Reports*, 372(1):1–129, December 2002.
- [86] Argyro Tasitsiomi, Andrey V. Kravtsov, Stefan Gottlöber, and Anatoly A. Klypin. Density profiles of Λ cdm clusters. *The Astrophysical Journal*, 607(1):125, may 2004.
- [87] J. Carretero, F. J. Castander, E. Gaztañaga, M. Crocce, and P. Fosalba. An algorithm to build mock galaxy catalogues using MICE simulations. *Monthly Notices of the Royal Astronomical Society*, 447(1):646–670, 12 2014.
- [88] Ramin A. Skibba and Ravi K. Sheth. A halo model of galaxy colours and clustering in the sloan digital sky survey. *Monthly Notices of the Royal Astronomical Society*, 392(3):1080–1091, January 2009.
- [89] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. A universal density profile from hierarchical clustering. *The Astrophysical Journal*, 490(2):493–508, December 1997.
- [90] O. et al Ilbert. Cosmos photometric redshifts with 30-bands for 2-deg2. *The Astrophysical Journal*, 690(2):1236–1249, December 2008.

- [91] G. Bruzual and S. Charlot. Stellar population synthesis at the resolution of 2003. *Monthly Notices of the Royal Astronomical Society*, 344(4):1000–1028, October 2003.
- [92] Robert C. Kennicutt. Star formation in galaxies along the hubble sequence. *Annual Review of Astronomy and Astrophysics*, 36(1):189–231, September 1998.
- [93] Mirko Curti, Filippo Mannucci, Giovanni Cresci, and Roberto Maiolino. The mass–metallicity and the fundamental metallicity relation revisited on a fully t_e-based abundance scale for galaxies. *Monthly Notices of the Royal Astronomical Society*, 491(1):944–964, November 2019.
- [94] B. P. et al. Abbott. A gravitational-wave standard siren measurement of the hubble constant. *Nature*, 551(7678):85–88, October 2017.
- [95] Michela Mapelli and Nicola Giacobbo. The cosmic merger rate of neutron stars and black holes. *Monthly Notices of the Royal Astronomical Society*, 479(4):4391–4398, June 2018.
- [96] Jose Maria Ezquiaga, Wayne Hu, Macarena Lagos, Meng-Xiang Lin, and Fei Xu. Modified gravitational wave propagation with higher modes and its degeneracies with lensing. *Journal of Cosmology and Astroparticle Physics*, 2022(08):016, August 2022.
- [97] Ignacio Magaña Hernandez and Anarya Ray. Beyond gaps and bumps: Spectral siren cosmology with non-parametric population models, 2024.
- [98] Piero Madau and Mark Dickinson. Cosmic star-formation history. *Annual Review of Astronomy and Astrophysics*, 52(1):415–486, August 2014.
- [99] Hebertt Leandro, Valerio Marra, and Riccardo Sturani. Measuring the Hubble constant with black sirens. *Phys. Rev. D*, 105(2):023523, 2022.
- [100] Matteo Martinelli, Francesca Scarcella, Natalie B. Hogg, Bradley J. Kavanagh, Daniele Gaggero, and Pierre Fleury. Dancing in the dark: detecting a population of distant primordial black holes, 2022.
- [101] Cristina Cigarrán Díaz and Suvodip Mukherjee. Mapping the cosmic expansion history from LIGO-virgo-KAGRA in synergy with DESI and SPHEREx. *Monthly Notices of the Royal Astronomical Society*, jan 2022.
- [102] Mohammadtaher Safarzadeh, Sylvia Biscoveanu, and Abraham Loeb. Constraining the delay time distribution of compact binary objects from the stochastic gravitational-wave background searches. *The Astrophysical Journal*, 901(2):137, oct 2020.
- [103] Jonathan R. Gair, Archisman Ghosh, Rachel Gray, Daniel E. Holz, Simone Mastrogiovanni, Suvodip Mukherjee, Antonella Palmese, Nicola Tamanini, Tessa Baker, Freija Beirnaert, Maciej Bilicki, Hsin-Yu Chen, Gergely Dályá, Jose Maria Ezquiaga, Will M. Farr, Maya Fishbach, Juan Garcia-Bellido,

- Tathagata Ghosh, Hsiang-Yu Huang, Christos Karathanasis, Konstantin Leyde, Ignacio Magaña Hernandez, Johannes Noller, Gregoire Pierra, Peter Raffai, Antonio Enea Romano, Monica Seglar-Arroyo, Danièle A. Steer, Cezary Turcki, Maria Paola Vaccaro, and Sergio Andrés Vallejo-Peña. The hitchhiker’s guide to the galaxy catalog approach for dark siren gravitational-wave cosmology. *The Astronomical Journal*, 166(1):22, June 2023.
- [104] Jiming Yu, Zhengyan Liu, Xiaohu Yang, Yu Wang, Pengjie Zhang, Xin Zhang, and Wen Zhao. Measuring the hubble constant with coalescences of binary neutron star and neutron star-black hole: bright sirens & dark sirens, 2023.
- [105] Hsin-Yu Chen, Maya Fishbach, and Daniel E. Holz. A two per cent hubble constant measurement from standard sirens within five years. *Nature*, 562(7728):545–547, October 2018.
- [106] Eric Thrane and Colm Talbot. An introduction to bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models. *Publications of the Astronomical Society of Australia*, 36, 2019.
- [107] Andreas Finke, Stefano Foffa, Francesco Iacovelli, Michele Maggiore, and Michele Mancarella. Cosmology with ligo/virgo dark sirens: Hubble parameter and modified gravitational wave propagation, 2021.
- [108] P. J. E. Peebles. *The large-scale structure of the universe*. 1980.
- [109] A. Palmese, R. Kaur, A. Hajela, R. Margutti, A. McDowell, and A. MacFadyen. A standard siren measurement of the hubble constant using gw170817 and the latest observations of the electromagnetic counterpart afterglow, 2023.
- [110] K. G. Malmquist. On some relations in stellar statistics. *Meddelanden fran Lunds Astronomiska Observatorium Serie I*, 100:1–52, March 1922.
- [111] D. Lynden-Bell, S. M. Faber, David Burstein, Roger L. Davies, Alan Dressler, R. J. Terlevich, and Gary Wegner. Photometry and Spectroscopy of Elliptical Galaxies. V. Galaxy Streaming toward the New Supergalactic Center. *apj*, 326:19, March 1988.
- [112] K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. Healpix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. *The Astrophysical Journal*, 622(2):759–771, April 2005.

Ringraziamenti

Come negli anni precedenti, continuare a lavorare nel mondo della ricerca mi ha mostrato ancora una volta come la scienza sia fatta di persone, una frase che sembra scontata, ma non è mai banale. Di questo ringrazio sentitamente il mio supervisor Enzo Branchini, per la sua innaturale pazienza nei miei confronti, la mia supervisor Melita Carbone e Marco Bonici, per il loro contributo scientifico e non.

Come ogni lavoro che ci porta a sporgersi in avanti, c'è bisogno di un appoggio solido che ci bilanci. Ringrazio dunque la mia famiglia e in particolare mia madre, per il lavoro svolto, sul quale, in ultima analisi, è stato possibile costruire il mio. Aggiungo poi Alice Merida, che come tutti i bambini, aiuta a mettere in prospettiva le cose. Spero un giorno tu abbia bisogno di dover attingere a questo testo.

Un doveroso e sentito ringraziamento va ai miei colleghi cosmologi, romani e non. Le missioni a Genova sono state un piacere anche e soprattutto per voi, e forse anche per l'asinello.

Ringrazio poi i dottorandi di Roma Tre, in particolar modo i pesciolini dell'acquario, passati presenti e perché no, pure quelli futuri. Lavorare in un ambiente così piacevole fa la differenza, e questo è stato possibile grazie a voi. In particolare devo ringraziare Sacha, compagno di scrivania e collega dalle "best vibes". Insegnarti a giocare a carte è stato un piacere. Un altro pesciolino a cui esprimere riconoscenza è Romana, grande sostegno fisico, morale e culinario, la cui pazienza e gentilezza hanno fatto la differenza.

Non è una tesi se non si menziona il gruppo CACCA, sempre lì a fare da mente alveare. Non so se traspare, ma colgo l'occasione per rendervi grazie, dell'aiuto pratico della compagnia e del fantamorto. Mettere in pausa le varie e accorate discussioni sui punteggi fino alla consegna di questo scritto è un piccolo gesto che nasconde molto affetto.

Ultimi, ma non per questo meno importanti i componenti di "*Inefficienti+Irene*", compreso il nuovo membro a Honorem Leonardo. Come detto anni fa, siete la famiglia che uno si sceglie e sono sempre più felice di rinnovare questa decisione ogni volta.