Stochastic gravitational wave background : detection methods in non-Gaussian regimes

Lionel Martellini

UNIVERSITÉ CÔTE D'AZUR

ARTEMIS Observatoire de la CÔTE d'AZUR

École doctorale Science Fondamentales et Appliquées

Unité de recherche UMR 6161

Thèse de doctorat

Présentée en vue de l'obtention du
grade de docteur en Astrophysique Relativiste
de
l'UNIVERSITE COTE D'AZUR

par

Lionel Martellini

Le Fond Gravitationnel Stochastique: Méthodes de Détection en Régimes Non-Gaussiens

Dirigée par Tania Régimbau

Soutenue le 23 Mai 2017
Devant le jury composé de :

Mme Chiara Caprini, chargée de recherche CNRS, APC	Examinatrice
M. Nelson Christensen, directeur de recherche CNRS, OCA	Examinateur
M. Jean-Daniel Fournier, directeur de recherche CNRS, OCA	Examinateur
Mme Tania Regimbau, chargée de recherche CNRS, OCA	Directrice de thèse
M. Joseph Romano, professeur, University of Texas Rio Grande Valley	Rapporteur
Mme Mairi Sakellariadou, professeur, King's College London	Rapporteure

École doctorale Science Fondamentales et Appliquées

Unité de recherche UMR 6161

Thèse de doctorat

Présentée en vue de l'obtention du
grade de docteur en Astrophysique Relativiste
de
l'UNIVERSITE COTE D'AZUR

par
Lionel Martellini

Stochastic Gravitational Wave Background: Detection Methods in Non-Gaussian Regimes

Dirigée par Tania Régimbau

Soutenue le 23 Mai 2017
Devant le jury composé de :

Mme Chiara Caprini, chargée de recherche CNRS, APC	Examinatrice
M. Nelson Christensen, directeur de recherche CNRS, OCA	Examinateur
M. Jean-Daniel Fournier, directeur de recherche CNRS, OCA	Examinateur
Mme Tania Regimbau, chargée de recherche CNRS, OCA	Directrice de thèse
M. Joseph Romano, professeur, University of Texas Rio Grande Valley	Rapporteur
Mme Mairi Sakellariadou, professeur, King's College London	Rapporteure

"Trop peu d'un firmament pour des millions d'astres"

Louis Aragon

Contents

Acknowledgments

First and foremost, I would like to express my gratitude to Tania Regimbau for starting a research collaboration with someone from a different academic field, and for making it work. Your continuous support over the last 5 years as a coauthor, as a PhD advisor, and also as a person, has truly meant a lot to me. I can only hope that we will maintain the same level and quality of interactions over the next 5 years and beyond. I also would like to thank Chiara Caprini, Nelson Christensen, Jean-Daniel Fournier, Joseph Romano and Mairi Sakellariadou for accepting to be members of my thesis committee. I feel fortunate to have such a distinguished group of astronomers and astrophysicists look over my work and provide comments and suggestions. A particularly large piece of gratitude goes to Nelson Christensen, who has offered extremely detailed feedback on a preliminary version of the document, and to Joe Romano, who has identified an inconsistency in the analysis presented in a previous version chapter 5 and has provided most helpful guidance in the revision process. Frans Pretorious and Paul Steinhardt, who have let me sit in their General Relativity and Cosmology PhD courses at Princeton University and gracefully answered my many naive questions along the way, have played an important role in my formal education in relativistic astrophysics, and for this I am grateful to them as well. On a different note, I would like to thank my father for having given me the taste for science, and my mother for having given me the taste for happiness. I finally would like to thank Daphne, Adhara, Calypso, Raphael and Theodore for their love, that mysterious form of energy that fuels every one of my steps.

Résumé

Les interféromètres de nouvelle génération devraient nous permettre de détecter le fond stochastique d'ondes gravitationnelles engendré par la superposition d'un nombre élevé de signaux gravitationnels aléatoires indépendants d'origine astrophysique ou cosmologique. La plupart des méthodes de détection du fond gravitationnel stochastique reposent sur l'hypothèse simplificatrice selon laquelle sa distribution ainsi que celle du bruit des détecteurs sont Gaussiennes. Le sujet principal de cette thèse est la mise en place de méthodes améliorées de détection du fond gravitationnel stochastique qui tiennent compte explicitement du caractère non-Gaussien de ces distributions. En utilisant un développement d'Edgeworth à l'ordre 4, nous obtenons dans un premier temps une expression analytique pour la statistique du rapport de vraisemblance en présence d'une distribution non-Gaussienne du fonds gravitationnel stochastique. Cette expression généralise l'expression habituelle lorsque la skewness (ou coefficient de symétrie) et l'excès de kurtosis (ou coefficient d'aplatissement) de la distribution du fond stochastique sont non nuls. Sur la base de simulations stochastiques pour différentes distributions symétriques présentant des queues plus épaisses que celles de la distribution Gaussienne, nous montrons par ailleurs que le 4eme cumulant peut-être estimé avec une précision acceptable lorsque le ratio signal à bruit est supérieur à 1%, ce qui devrait permettre d'apporter des contraintes supplémentaires intéressantes sur les valeurs de paramètres issus des modèles astrophysiques et cosmologiques. Dans un deuxième temps, nous cherchons à analyser l'impact sur les méthodes de détection du fond gravitationnel stochastique de déviations par rapport à la normalité dans la distribution du bruit des détecteurs. Pour des valeurs raisonnables des paramètres, nous montrons que tenir compte explicitement de la non-normalité de la distribution du bruit a un impact substantiel sur les méthodes de détection, et conduit à des estimations plus élevées des probabilités de non-détection pour des niveaux donnés de probabilités de fausse alarme.

Summary

The new generation of interferometers should allow us to detect stochastic gravitational wave backgrounds that are expected to arise from a large number of random, independent, unresolved events of astrophysical or cosmological origin. Most detection methods for gravitational waves are based upon the assumption of Gaussian gravitational wave stochastic background signals and noise processes. Our main objective is to improve the methods that can be used to detect gravitational backgrounds in the presence of non-Gaussian distributions. We first maintain the assumption of Gaussian noise distributions so as to better focus on the impact of deviations from normality of the signal distribution in the context of the standard cross-correlation detection statistic. Using a 4th-order Edgeworth expansion of the unknown density for the signal and noise distributions, we first derive an explicit expression for the non-Gaussian likelihood ratio statistic, which is obtained as a function of the variance, but also skewness and kurtosis of the unknown signal and noise distributions. We use numerical procedures to generate maximum likelihood estimates for the gravitational wave distribution parameters for a set of symmetric heavy-tailed distributions, and we find that the fourth cumulant can be estimated with reasonable precision when the ratio between the signal and the noise variances is larger than 1%, which should be useful for analyzing the constraints on astrophysical and cosmological models. In a second step, we analyze the efficiency of the standard cross-correlation statistic in situations that also involve non-Gaussian noise distributions. For reasonable parameter values, we find that properly accounting for the presence of non-Gaussian distributions as opposed to wrongly assuming that higher-order cumulants of the noise distributions are zero has material implications in the implementation of standard detection procedures in that it generates substantially higher values for probabilities of false dismissal corresponding to given levels of probabilities of false alarm.

Résumé Substantiel

Les ondes gravitationnelles (ou OGs) sont des perturbations de la métrique de l'espace-temps se propageant à la vitesse de la lumière, engendrées par une modification brutale et asymétrique du contenu de masse-énergie en un point de l'univers, et dont l'existence a été prédite en 1916 par Albert Einstein [73] comme conséquence de la théorie de la relativité générale qu'il avait introduite à l'occasion de 4 papiers publiés en Novembre 1915 ([72], [67], [71] and [66]). La première confirmation observationnelle de l'existence des OGs fut apportée par Hulse et Taylor en 1975 par l'observation du système binaire PSR1913+16 [98], et l'analyse de la perte d'énergie de ce système par Taylor et Weisberg en 1982 [164] (voir également [165], [177], et [176] pour des analyses plus récentes du système PSR1913+16). Ces travaux ont montré que la période orbitale du système binaire PSR1913+16 constitué de deux étoiles à neutrons décroissait d'un millième de seconde par an, une mesure en accord avec la prévision théorique concernant l'émission d'ondes gravitationnelles pour un tel système. A la suite de cette détection indirecte, la détection directe d'un signal d'OG est devenue une question centrale en astrophysique relativiste, et d'importantes ressources ont été consacrées à la mise au point d'interféromètres permettant de détecter ces OGs. Le 14 Septembre 2015, la nouvelle génération des détecteurs LIGO, situés respectivement à Hanford dans l'Etat de Washington et à Livingston dans l'Etat de Louisiane, a permis de détecter un signal gravitationnel [9]. L'analyse de l'amplitude des ondes et de l'évolution de leur fréquence a révélé qu'elles avaient été produites par la coalescence d'un système binaire composé de deux trous noirs situés à environ 410 Megaparsec de notre galaxie, respectivement de 29 et de 36 fois la masse du Soleil. Cette fusion a engendré un trou noir de 62 fois la masse du Soleil, les 3 masses solaires manquantes ayant été dissipées sous forme d'ondes gravitationnelles. Une deuxième détection a eu lieu le 26 Décembre 2015, qui a porté à nouveau sur la coalescence d'un système binaire composé de deux trous noirs situés à environ 440 Megaparsec de notre galaxie, respectivement de 14.2 et de 7.5 masses solaires. Ces récentes détections, qui suggèrent que la population des binaires de trous noirs est

peut-être plus abondante qu'initialement envisagé mais surtout que la masse de ces trous noirs est peut-être plus élevée que prévu, ont ouvert une ère nouvelle en astronomie, celle de l'astronomie gravitationnelle, et l'on s'attend désormais à ce qu'une troisième génération de détecteurs (projets Einstein Telescope [142], LIGO Voyager ou Cosmic Explorer [124]) permette d'augmenter la probabilité de détecter des OGs de magnitude encore plus faible grâce à des gains de sensibilité supplémentaires.

Au-delà des signaux gravitationnels individuels résolus comme GW150914 et GW151226, les interféromètres de nouvelle génération devraient nous permettre de détecter le fond stochastique d'OGs engendré par l'addition d'un nombre élevé de signaux gravitationnels aléatoires indépendants non résolus d'origine astrophysique ou cosmologique. La stratégie optimale de détection du fond gravitationnel stochastique consiste à prendre le produit croisé des détections d'au moins deux détecteurs afin d'éliminer au mieux le bruit de l'instrument [15]. Dans le cadre d'hypothèses classiques incluant la présence de distributions stationnaires et de corrélation temporelle nulle pour les signaux stochastiques et le bruit, ainsi que la présence de détecteurs de même sensibilité, il est possible de montrer que cette statistique de détection dite de *cross-correlation* (CC) est optimale au sens où elle permet de minimiser la probabilité de rejeter à tort l'hypothèse d'une détection pour un taux donné de fausse alarme (voir par exemple [44, 78, 15]). Les méthodes standard de détection des OGs reposent par ailleurs sur l'hypothèse simplificatrice selon laquelle le fond gravitationnel stochastique est distribué de façon Gaussienne. Des travaux récents portant sur une modélisation réaliste de la population des objets astrophysiques de nature à engendrer des OGs d'amplitude assez forte pour être détectées laisse cependant supposer qu'il n'existe pas un nombre de sources superposées assez élevé pour permettre l'application du théorème central limite, et que le fond stochastique résultant de la superposition de ces signaux aléatoire peut donc faire apparaître des déviations par rapport à l'hypothèse de normalité. Il a également été montré que le fond stochastique engendré par des cordes cosmiques pourrait être dominé par la contribution non-Gaussienne des sources les plus proches [55, 144].

Le sujet principal de cette étude est la mise en place de méthodes de détection du fond gravitationnel stochastique qui tiennent compte explicitement du caractère non-Gaussien de sa distribution, ce qui devrait permettre d'améliorer l'efficacité de la procédure de détection, et également de potentiellement permettre de distinguer le fond d'origine astrophysique du fond d'origine cosmologique. L'approche que nous proposons dans un premier temps est basée sur le développement d'Edgeworth, qui est un développement formel en série infinie permettant d'écrire la fonction caractéristique d'une distribution non-Gaussienne inconnue comme une perturbation autour de la fonction caractéristique de la distribution Gaussienne. Dans la mesure où le développement d'Edgeworth a précisément vocation à caractériser la déviation à la normalité dans le cadre d'une application à distance finie du théorème central limite, il apparaît naturellement adapté à l'analyse du fond gravitationnel stochastique résultant de la superposition d'un nombre fini de sources d'origine astrophysique, sous réserve que la non-Gaussianité ainsi engendrée ne soit pas trop forte. En utilisant un développement d'Edgeworth poussé jusqu'à l'ordre 4, nous avons réussi à obtenir une expression analytique pour la statistique de détection dans un cadre non-Gaussien qui fait intervenir non seulement la variance mais aussi le coefficient d'asymétrie (skewness) et le coefficient d'aplatissement (kurtosis) de la distribution du signal. Cette expression généralise l'expression habituelle, qui est obtenue comme cas particulier de l'expression générale lorsque les cumulants d'ordre 3 et 4 de la distribution du fond stochastique sont nuls, comme c'est le cas pour une distribution Gaussienne. Sur la base de simulations stochastiques pour différentes distributions symétriques présentant des queues plus épaisses que celles de la distribution Gaussienne, nous montrons par ailleurs que le 4eme cumulant peut-être estimé avec une précision acceptable lorsque le ratio signal à bruit est supérieur à 1%. Cette valeur de 1% est à comparer à la valeur attendue du rapport signal à bruit pour la détection du fond gravitationnel stochastique par la deuxième génération des détecteurs LIGO et Virgo (Advanced LIGO et Advanced Virgo) et par Einstein Telescope, valeur qui est comprise entre 1% et 10%. Pour les cordes cosmiques, la valeur de la densité d'energie à une fréquence de 100 Hz est attendue entre

les valeurs 10^{-9} et 10^{-5}, ce qui correspond à un rapport signal à bruit de 10^{-6} à 1 pour la deuxième génération des détecteurs LIGO et Virgo et de $10^{-4} - 100$ pour Einstein Telescope. Pour les fusions d'objets compacts, la valeur de la densité d'energie à une fréquence de 100 Hz est attendue entre les valeurs 10^{-10} et 10^{-7}, ce qui correspond à un rapport signal à bruit de 10^{-7} à 0.01 pour la deuxième génération des détecteurs LIGO et Virgo et de 10^{-5} à 1 pour Einstein Telescope.

L'intérêt principal de la méthode est précisément de permettre l'estimation de paramètres supplémentaires, à savoir les cumulants d'ordre 3 et 4 de la distribution du signal gravitationnel, ce qui devrait permettre d'apporter des contraintes supplémentaires intéressantes sur les valeurs de paramètres issus des modèles astrophysiques et cosmologiques. Il s'avère par exemple impossible de distinguer, dans le cadre des méthodes d'estimation traditionnelles de type *cross-correlation* (CC), un fond stochastique provenant de systèmes de binaires compactes (étoiles à neutron et/ou trous noirs) caractérisés par un taux de coalescence élevé et des masses faibles, ou bien un taux de coalescence faible et des masses élevées car ces deux situations peuvent donner un signal de même amplitude [126]. La méthode introduite ici pourra en principe permettre de distinguer ces deux situations très différentes sur le plan astrophysique par l'estimation du cumulant d'ordre 4, sachant que le premier cas (taux élevé et masses faibles) correspond à un signal de type continu et une valeur faible de la kurtosis tandis que le deuxième cas (taux faible et masses élevées) correspond au contraire a un signal de type "popcorn" avec une valeur élevée de la kurtosis. Au final, l'un des avantages principaux de l'approche proposée est son caractère non-paramétrique, qui permet de s'affranchir de la nécessite de faire des hypothèses restrictives à propos de la nature exacte de la distribution non-Gaussienne sous-jacente. Les méthodes développées dans le cadre de ces travaux pourraient a priori être utilisées dans le cadre d'un effort de distinction du fond d'origine astrophysique et du fond d'origine cosmologique, sous réserve que les déviations à la normalité admettent dans ces deux cas des signatures bien distinctes.

Dans cette approche, nous avons maintenu l'hypothèse d'une distribution Gaussienne

du bruit instrumental et environnemental de façon à mieux isoler l'impact d'une déviation de l'hypothèse de normalité pour la distribution du signal. Il existe cependant de bonnes raisons de penser que le bruit instrumental ou environnemental est également distribué de manière non-Gaussienne [181], et il est alors utile de chercher à analyser l'impact sur les méthodes de détection du fond gravitationnel stochastique de ces déviations par rapport à la normalité dans la distribution du bruit. Par ailleurs, l'hypothèse standard selon laquelle les détecteurs auraient la même sensibilité pourrait ne pas être strictement vérifiée lors des étapes intermédiaires de calibration des détecteurs de deuxième génération Advanced LIGO and Advanced Virgo, et une telle hypothèse serait encore plus discutable dans le cas d'observations jointes impliquant des détecteurs de deuxième et troisième générations. Dans ce contexte, nous cherchons à analyser la performance de la statistique CC standard dans des situations impliquant des déviations des hypothèses évoquées ci-dessus, et en particulier des situations impliquant la présence de non-normalité dans la distribution du bruit des détecteurs.

Pour cela, nous montrons d'abord qu'il est possible d'obtenir une expression généralisée pour la statistique du rapport de vraisemblance dans un cadre de travail impliquant une déviation de la normalité non seulement pour le signal mais aussi pour le bruit des détecteurs. Cette expression analytique permet d'envisager une estimation efficace des paramètres de variance, skewness et kurtosis de la distribution du signal et du bruit des détecteurs. En parallèle a l'identification des estimateurs du maximum de vraisemblance, nous introduisons également des estimateurs basés sur la méthode dite des moments, qui sont sans biais par construction. Pour des valeurs raisonnables des paramètres, nous montrons que tenir compte explicitement de la non-normalité de la distribution du bruit a un impact substantiel sur les méthodes de détection. Nous introduisons en particulier une expression analytique pour l'espérance et la variance de la statistique de détection standard dans un cadre non-Gaussien, et nous montrons pour des valeurs raisonnables des paramètres que tenir compte explicitement de la non-normalité de la distribution du bruit a un impact substantiel sur les méthodes de détection, et conduit à des estimations

plus élevées des probabilités de non-détection pour des niveaux donnés de probabilités de fausse alarme. Au-delà de leur impact sur la méthodes de détection CC standard, nos résultats suggèrent également qu'il est possible d'obtenir une statistique de détection optimale dans un cadre non-Gaussien généralisé. Cette statistique de détection généralise la statistique de détection standard, qui est recouverte pour des valeurs nulles des cumulants d'ordre 3 et 4 des distributions du signal et du bruit.

1

Introduction

Gravitational waves (GWs in short) are perturbations of spacetime geometry travelling at the speed of light created by asymmetric acceleration of masses or non stationary fields, which existence is predicted by the theory of general relativity (GR in short). While electromagnetic waves interact strongly with matter, GWs minimally interacting with the matter they encounter, which allows us to probe astrophysical or cosmological phenomena that cannot be observed by electromagnetic signals, such as the inspiral, coalescence and merger of black holes, the collapse of a stellar core, or the dynamics of the early Universe. The first observational validation of the existence of gravitational waves is the PSR B1913+16 binary pulsar system discovered by Hulse and Taylor in 1975 [98], and for which Taylor and Weisberg [164] subsequently demonstrated that the rate of decay of the orbit exactly matches GR predictions regarding the loss of energy of the system due to the emission of gravitational waves (see also [165], [177], and [176] for more recent observations and related analyses). Following this first indirect evidence of the existence of GWs, the direct detection of GWs has become a question of central importance in relativistic astrophysics, and an increasing range of efforts has been dedicated to the design of improved detectors. On September 14, 2015, second generation Laser Interferometer Gravitational-Wave Observatory detectors (known as Advanced LIGO detectors),

located in Hanford, Washington, and Livingston, Louisiana, USA, have successfully detected GWs produced during the late inspiral and merger of two black holes of masses respectively 29 and 36 solar masses, located at approximately 410 Megaparsec from our galaxy [9]. The black hole resulting from the merger had a total mass of 62 solar masses, with the missing 3 solar masses having been carried away under the form of gravitational waves. A second detection of gravitational waves generated by the coalescence of a binary system of stellar mass black holes subsequently took place on December 26, 2015 [5]. These remarkable detections have opened a new era of astronomy, and third-generation interferometers such as Einstein Telescope [142], LIGO Voyager or Cosmic Explorer [124] are expected to further increase the likelihood of detecting the exceedingly small effects of gravitational waves.

In addition to *resolved* individual GW signals such as GW150914 and GW151226, the new generations of interferometers should allow us to detect stochastic GW *backgrounds*, which are expected to arise from a large number of random, independent, unresolved events of astrophysical or cosmological origin. The optimal detection strategy to search for a stochastic background is to cross correlate the output of two detectors (or of a network of detectors) to eliminate the instrumental noise [15]. Under standard assumptions including stationary and serially uncorrelated Gaussian gravitational wave stochastic background signal and noise processes as well as homogenous detector sensitivities, the standard cross-correlation (CC in short) detection statistic is known to be optimal in the sense of minimizing the probability of a false dismissal at a fixed value of the probability of a false alarm (see for example [44, 78, 15]). While the GW background is usually assumed to be Gaussian invoking the central limit theorem, and thus completely characterized by its mean and variance, recent predictions based on population modeling suggest that for many astrophysical models, there may not be enough overlapping sources, resulting in the formation of a non-Gaussian background. [182] show that the population of binary black hole systems in the observable universe could produce for reasonable mass and rate parameter values a series of non-continuous background burst signals that will most likely

be of the shot noise or at most popcorn noise type (see section 3.1 for precise definitions) in the Advanced LIGO/Virgo frequency range, while the signal will be continuous but not necessarily Gaussian for ET type-detectors depending upon assumed parameter values. Turning to cosmological sources of gravitational waves, it has also been shown that the background from cosmic strings could be strongly non-Gaussian [55, 144] (see section 3.1.2 for more details).

Our work has a main focus on improving the methods that can be used to detect non-Gaussian GW backgrounds, which would permit to possibly distinguish between astrophysical and cosmological GW backgrounds and gain confidence in a detection. The approach we propose in a first step is based on Edgeworth expansion, which is a formal asymptotic expansion of the characteristic function of the signal distribution, whose unknown probability density function is to be approximated in terms of the characteristic function of the Gaussian distribution. Since the Edgeworth expansion provides asymptotic correction terms to the Central Limit Theorem up to an order that depends on the number of moments available, it is ideally suited for the analysis of stochastic gravitational wave backgrounds generated by a finite number of astrophysical sources. It is also well-suited for the analysis of signals from cosmological origin in case the deviations from the Gaussian assumption are not too strong. Using a 4th-order Edgeworth expansion, we obtain an explicit expression for the nearly optimal non-Gaussian likelihood statistic that is obtained as a function of the variance, but also skewness and kurtosis, of the unknown signal distribution. This expression generalizes the standard maximum likelihood detection statistic, which is recovered in the limit of vanishing third and fourth cumulants of the empirical conditional distribution of the detector measurement. We use numerical procedures to generate maximum likelihood estimates for the gravitational wave distribution parameters for a set of symmetric heavy-tailed distributions, and we find that the fourth cumulant can be estimated with reasonable precision when the ratio between the signal and the noise variances is larger than 1%. The main benefit of the procedure is precisely that it allows us to estimate additional parameters, namely the 3rd and 4th

cumulants of the gravitational wave signal distribution, which should be useful for analyzing the constraints on astrophysical and cosmological models that will be imposed by observed gravitational wave signals, and comparing them to the constraints derived from supernovae or galaxy clusters observations. Overall, one key advantage of the proposed methodology, which relies on an explicit correction to the central limit theorem when the number of sources is finite, is that can be applied without any assumption regarding the exact nature of the departure from normality.

In this first step, we maintain the assumption of Gaussian noise distributions so as to better focus on the impact of deviations from normality of the signal distribution in the context of the standard cross-correlation detection statistic. This assumption is at odds, however, with accumulated evidence of strong deviations from the Gaussian assumption for noise distributions in gravitational waves detectors [125]. A recent paper [181] introduces a new measure for characterizing the non-Gaussian noise component modelled as a Student-t distribution and reveals stationary and transient deviations from Gaussianity in LIGO S5 data. This is a serious concern since existing detection strategies for both deterministic and stochastic signals are expected to deteriorate when non-Gaussian noise is present [13, 14]. If the exact non-Gaussian nature of the detector noise is understood, it is possible to introduce a robust detection statistic using the specific non-Gaussian noise assumption (see for example [16] for a detection method based on an exponentially distributed noise process, and [152] for a detection method based on a Student's t distributed noise process). Given that the actual noise distribution is a priori unknown, it is unclear how much improvement, if any, these methods would allow with respect to the standard method based upon a Gaussian assumption in case of a misspecification of the exact deviation from the Gaussian assumption. In this context, we analyze the efficiency of the CC statistic in situations that deviate from the Gaussian assumption for both the stochastic gravitational wave signal distribution and detector noise distributions. To do so we first show how to obtain consistent estimates for the first four cumulants of the signal and noise distributions using a suitable extension of the likelihood function, for

which we derive an analytical expression. These results extend our previous results where we have Focused on a situation involving a non-Gaussian signal distribution, but have maintained the assumption of a Gaussian noise distribution. In addition to obtaining parameter estimates through maximum likelihood techniques, we also introduce so-called moment-based estimators given by analytical functions of the joint observations from the two detectors. While the moment-based estimators for the variance of the signal and the noise in each detector coincide with the maximum likelihood estimators in the Gaussian case, the moment-based estimators may be different in a generalized non-Gaussian setting but they share with maximum likelihood estimators the desirable property to be unbiased by construction. Turning to a numerical analysis, we find that properly accounting for the presence of non-Gaussian distributions as opposed to wrongly assuming that higher-order cumulants of the noise distributions are zero has material implications in the implementation of standard detection procedures in that it generates higher values for probabilities of false dismissal corresponding to given levels of probabilities of false alarm. The correction is found to be particularly substantial when detector sensitivities exhibit substantial differences, a situation that is expected to hold in early phases of development of the Advanced LIGO-Virgo detectors before they reach their design sensitivity. Under such circumstances, or in joint detections from Advanced LIGO and the Einstein Telescope project [142], failing to account for the presence of non-Gaussian detector noise distributions In addition to their implications for the performance of the standard CC detection statistic, we also discuss the implications of our results for the derivation of an optimal detection statistic in a non-Gaussian context.

The rest of this thesis is organized as follows. In chapter 2, we provide a broad introduction to the theory of general relativity and its implications for the generation of gravitational waves. In chapter 3, we propose an analysis of the stochastic gravitational wave background and its distribution. In chapter 4, we introduce a semi-parametric approach to the detection of non-Gaussian stochastic gravitational wave backgrounds. In chapter 5, we analyze the efficiency of the standard cross-correlation statistic in the

presence of non-Gaussian detector noise distributions and discuss the implications for the derivation of an optimal detection statistic in a non-Gaussian setting. Finally, we present in chapter 6 our conclusions and suggestions for further research. Note that we have chosen to write the chapters so that they can be read somewhat independently, even if this necessarily implies some redundancies.

2

Gravitational Waves: Theory, Detection and Sources

We first provide a brief introduction to general relativity and gravitational waves, before presenting an overview of gravitational wave detectors as well as an overview of the main sources of gravitational waves.

2.1 Introduction to General Relativity, Cosmology and Gravitational Waves

General relativity (or GR in short) is a theory of gravitation introduced by Albert Einstein in 4 papers published in November 1915 ([72], [67], [71] and [66]). It has become the commonly accepted description of gravitation in modern physics, which provides a unified description of gravity as a geometric property of spacetime.

2.1.1 Introduction to General Relativity

The presentation that follows is broadly inspired by [39]. In what follows we adopt the usual notational convention for indices, namely greek letters $\alpha, \beta, \gamma...$ are used for 4-dimensional spacetime indices ranging from 0 (the time coordinate) to 3 (for the three space dimensions), while roman letters $i, j, k, ...$ are used for 3-dimensional spatial indices. Throughout the text, we also use the Einstein summation convention regarding the repeated adjacent indices in upwards and downwards location. In other words, we take:

$$u^{\alpha}v_{\alpha} \equiv \sum_{\alpha=0}^{3} u^{\alpha}v_{\alpha}. \tag{2.1}$$

The Special Theory of Relativity (SR): Principle of Relativity and Spacetime Geometry with Inertial Reference Frames

The *special theory of relativity* published in the "annus mirabilis" 1905 ([70], [69]) by Albert Einstein is based upon the assumption that there is no preferred inertial reference frame, or in other words that measurements of physical quantities and expressions of physical laws remain the same after changing from a reference frame to another reference frame that is in constant rectilinear motion with respect to the original one. In particular, this postulate implies that the speed of light will yield the same measure $c = 299,792,458$ ms^{-1} in all referential frames. The geometric structure of the spacetime used in special relativity is the so-called Minkowski (or Poincaré-Minkowski) spacetime, a generalization of the Euclidean space where the squared distance ds^2 between the point $P = (x_1, x_2, x_3)$ (we sometimes also use the notation x, y, z for the 3 spatial coordinates) and an infinitesimally close point $Q = (x_1 + dx_1, x_2 + dx_2, x_3 + dx_3)$ is given by Pythagorean theorem as:

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2. \tag{2.2}$$

Minkowski spacetime is a flat spacetime where physical events are described in terms of spacetime coordinates (t, x_1, x_2, x_3) and where the infinitesimal spacetime interval between nearby events is measured in terms of the line element ds^2 given in Cartesian coordinates by the following generalization of Pythagorean theorem:

$$ds^2 = -c^2dt^2 + dx_1^2 + dx_2^2 + dx_3^2, \tag{2.3}$$

or equivalently by:

$$ds^2 = -dx_0^2 + dx_1^2 + dx_2^2 + dx_3^2, \tag{2.4}$$

in the coordinate system $(x_0 = ct, x_1, x_2, x_3)$. Defining the Minkowski *metric* $\eta_{\alpha\beta}$ by:

$$\eta_{00} = -1 \tag{2.5}$$

$$\eta_{ii} = 1 \text{ for } i = 1, 2, 3 \tag{2.6}$$

$$\eta_{\alpha\beta} = 0 \text{ for } \alpha \neq \beta, \tag{2.7}$$

we can write the line element ds^2 in Minkowski spacetime as:

$$ds^2 = \eta_{\alpha\beta}dx^\alpha dx^\beta. \tag{2.8}$$

In matrix notation, the metric is simply:

$$\eta = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.9}$$

Equation 2.8 defines the geometry of Minkowski spacetime. The symmetry group of this geometry is the group of coordinate transformations $\Lambda^\mu_\nu : (x_0, x_1, x_2, x_3) \rightarrow (x'_0, x'_1, x'_2, x'_3)$ that leaves the quadratic form 2.8 of the interval ds^2 invariant. This group of coordi-

nate transformations, known as the Lorentz group, is therefore defined by the following transformations:

$$x'^{\mu} = \Lambda^{\mu}_{\nu} x^{\nu} + a^{\mu}, \tag{2.10}$$

where the so-called *Lorentz boost* transformation Λ^{μ}_{ν} satisfies the condition:

$$\eta_{\alpha\beta} = \Lambda^{\alpha}_{\mu} \Lambda^{\beta}_{\nu} \eta_{\mu\nu} \tag{2.11}$$

so as to ensure the invariance of the line element 2.8. More generally, the *principle of relativity*, which requires that the laws of physics have the same expression in all inertial reference frames, implies that physical theories should be invariant under a class of transformations known as Poincaré transformations, which includes Lorentz boost transformations, but also translations and rotations. Physically, consider an observer in an initial frame of reference frame with coordinates (t, x_1, x_2, x_3) and an observer in a different frame of reference with coordinates (t', x'_1, x'_2, x'_3) moving with constant velocity v along the x_1-axis. The change in coordinates that leaves the spacetime interval invariant $ds' = ds$ is given by the following Lorentz-boost transformation along the x_1-axis:

$$\begin{cases} t' = \gamma \left(t - \frac{v}{c^2} x_1 \right) \\ x'_1 = \gamma \left(x_1 - vt \right) \\ x'_2 = x_2 \\ x'_3 = x_3 \end{cases} \tag{2.12}$$

where the so-called *Lorentz factor* is given by:

$$\gamma \equiv \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}. \tag{2.13}$$

Note that for a particle with fixed space coordinates ($dx_1^2 = dx_2^2 = dx_3^2 = 0$) the interval of time 2.8 elapsed as time moves forward ($dt > 0$) is negative:

$$ds^2 = -c^2 dt^2 < 0. \tag{2.14}$$

This leads us to the introduction of the *proper time* τ via

$$d\tau^2 = dt^2 = -\frac{ds^2}{c^2}, \tag{2.15}$$

or equivalently:

$$d\tau = \frac{1}{c}\sqrt{-ds^2}. \tag{2.16}$$

Proper time elapsed along a trajectory through spacetime parametrized a a function of some parameter $x^\alpha(\lambda)$ is thus defined as the time measured by a clock following that line, in the reference frame where the spatial coordinates do not vary (an observer on a different frame of reference will measure a different time). Integrating 2.16, we obtain that the proper time *interval* between two events on a trajectory is:

$$\Delta\tau = \int \frac{1}{c}\sqrt{-ds^2} = \int \frac{1}{c}\sqrt{-\eta_{\alpha\beta}dx^\alpha dx^\beta} \tag{2.17}$$

If the trajectory through spacetime is parametrized a a function of some parameter $x^\alpha(\lambda)$, then the proper time can be expressed as:

$$\Delta\tau = \int \frac{1}{c}\sqrt{-\eta_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda}}d\lambda \tag{2.18}$$

We thus have from 2.17:

$$\tau = \int \sqrt{dt^2 - \frac{1}{c^2}\left(dx_1^2 + dx_2^2 + dx_3^2\right)} \tag{2.19}$$

$$= \int \sqrt{1 - \frac{1}{c^2}\left[\left(\frac{dx_1}{dt}\right)^2 + \left(\frac{dx_2}{dt}\right)^2 + \left(\frac{dx_3}{dt}\right)^2\right]}dt \tag{2.20}$$

$$= \int \sqrt{1 - \frac{v^2(t)}{c^2}}dt = \int \frac{dt}{\gamma(t)} \tag{2.21}$$

where we have introduced the following quantity:

$$\gamma(t) \equiv \frac{1}{\sqrt{1 - \frac{v^2(t)}{c^2}}}, \tag{2.22}$$

which generalizes the Lorentz factor 2.13 to the case of a possibly time-varying velocity. The corresponding tangent vector

$$U^\mu = \frac{dx^\mu}{d\tau} \tag{2.23}$$

is called the *four-velocity* and is is automatically normalized:

$$\eta_{\mu\nu}U^\mu U^\nu = -1. \tag{2.24}$$

A related quantity is the *momentum four-vector* defined by:

$$p^\mu = mU^\mu, \tag{2.25}$$

where m is the *rest mass* of the particle. We can also define the *proper distance* as the distance between the two events as measured in an inertial frame of reference in which the events are simultaneous. In the same spirit, the proper length or rest length L_0 of an object is the length of the object measured by an observer which is at rest relative to it, by applying standard measuring rods on the object. The measurement of the object endpoints does not have to be simultaneous, since the endpoints are constantly at rest

at the same positions in the object rest frame, so it is independent of Δt. However, in relatively moving frames the object endpoints have to be measured simultaneously, since they are constantly changing their position.[1]

The General Theory of Relativity (GR): Principles of Equivalence and General Covariance and Spacetime Geometry with Accelerated Reference Frames

The *principle of relativity*, which is the founding principle of special relativity, has been extended by Einstein to the *principle of general covariance*, which requires the invariance of the form of physical laws under changes of reference frames that extend beyond the inertial reference frames, and which can include *accelerated* reference frames. The starting point was the recognition in 1907 [68] by Albert Einstein that the equivalence between inertial mass and gravitational mass, which implies the universality of free fall (initially noted by Galilée in 1638 century, and subsequently confirmed by Newton in 1687 and von Eötvös at the end of the 19th century), can be translated as what is now known as the *weak form of the equivalence principle*. Based on free-fall thought experiments, this principle states the equivalence between gravitation and acceleration: gravity can be canceled (free fall) or mimicked (constant acceleration) by acceleration.

Such general accelerated reference frames are mathematically represented by arbitrary differentiable coordinate transformations. The transformation from an inertial reference frame (x_0, x_1, x_2, x_3) to a general non inertial reference frame (x'_0, x'_1, x'_2, x'_3) is a non-linear transformation defined through the 4 functions $x'^{\alpha}(x^{\mu})$ expressing the new primed coordinates in terms of the original unprimed coordinates assumed to be expressed in an inertial frame of reference, or equivalently through the reverse transformation $x^{\mu}(x'^{\alpha})$ expressing the original unprimed coordinates as functions of the new primed coordinates. The non-linearity of the transformation implies that the line element ds^2 will take a more complicated form in the accelerated frame of reference compared to the form given in

[1] For this reason, the distance defined in equation 2.166 is not a proper distance. Indeed, it is measured between two free-moving test particles subject to gravitational wave oscillations A and B as $\frac{c}{2}(t_2 - t_1)$, where t_1 is the proper time for an observer located on a reference frame attached to A who sends a light signal in direction to B, which is reflected back to A where it arrives at the proper time $t_2 > t_1$.

equation 2.8 in the inertial frame of reference. Under this general transformation, and given that

$$dx^{\mu} = \frac{\partial x^{\mu}}{\partial x'^{\alpha}} dx'^{\alpha}, \tag{2.26}$$

starting with the line element given by special relativity in the inertial reference frame (x_0, x_1, x_2, x_3)

$$ds^2 = \eta_{\alpha\beta} dx^{\alpha} dx^{\beta}, \tag{2.27}$$

we obtain in the accelerated reference frame (x'_0, x'_1, x'_2, x'_3):

$$ds^2 = g_{\alpha\beta} dx'^{\alpha} dx'^{\beta}, \tag{2.28}$$

with

$$g_{\alpha\beta} = \eta_{\mu\nu} \frac{\partial x^{\mu}}{\partial x'^{\alpha}} \frac{\partial x^{\nu}}{\partial x'^{\beta}}.$$

Given the non-linearity of the 4 functions $x'^{\alpha}(x^{\mu})$, the functions $g_{\alpha\beta}$ exhibit in general an explicit dependence on the coordinates x^{μ}. As a result, the local geometry of spacetime is no longer given by the simple Minkowski metric in equation 2.8 with constant coefficients, but by the much more general quadratic metric in equation 2.28.[2] In this general spacetime endowed with the metric $g_{\mu\nu}$, the invariance of the line element implies

$$ds^2 = g_{\mu\nu} dx^{\mu} dx^{\nu} = g_{\mu\nu} \frac{\partial x^{\mu}}{\partial x'^{\alpha}} \frac{\partial x^{\nu}}{\partial x'^{\beta}} dx'^{\alpha} dx'^{\beta} = g'_{\alpha\beta} dx'^{\alpha} dx'^{\beta}, \tag{2.29}$$

with:

$$g'_{\alpha\beta} = g_{\mu\nu} \frac{\partial x^{\mu}}{\partial x'^{\alpha}} \frac{\partial x^{\nu}}{\partial x'^{\beta}}. \tag{2.30}$$

Such general metric spaces have been introduced by the mathematicians Gauss and Riemann in the 19th century in the situation where the quadratic form is positive definite.[3] More formally, let M be a $n-$dimensional C^{∞} manifold, $T_x M$ the tangent space at

[2] The geometry is still flat, and the departure from the Minkowski merely reflects a change of coordinates.

[3] In special and general relativity, the quadratic form is not definite positive and the metric is known

$x \in M$ and $TM \equiv \cup_{x \in M} T_x M$ the tangent bundle of M. Hence, each element of TM has the form (x, y), where $x \in M$ and $y \in T_x M$. In Riemannian geometry, the useful information about the curvature of a manifold is contained in the metric. We first introduce a function F, known as *metric function* or *generator function*, which measures the distance between two points $x = (x^1, x^2, ..., x^n)$ and $x + dx = F\left(x^1 + dx^1, x^2 + dx^2, ..., x^n + dx^n\right)$:

$$ds = F\left(x^1, x^2, ..., x^n, dx^1, dx^2, ..., dx^n\right). \tag{2.31}$$

Here $(x^1, x^2, ..., x^n)$ are the coordinates assigned in a given coordinate system to point x of M, and $(dx^1, dx^2, ..., dx^n)$ or $(y^1, y^2, ..., y^n)$ are coordinates of $y \in T_x M$, defined through the natural basis $e_i = \frac{\partial}{\partial x^i}\Big|_x$, with $y = y^i e_i$. Natural conditions (which may not be necessary for some/all physical applications) that should be satisfied by the function $F\left(x^1, ..., x^n, y^1, ..., y^n\right)$, denoted by $F(x, y)$, are as follows.

1. Positivity: $F(x, y) > 0$ for any $y = dx$.

2. Positive homogeneity: $F(x, py) = pF(x, y)$ for any $p > 0$ (F is a homogenous function of degree 1).[4]

3. Symmetry: $F(x, -y) = F(x, y)$. (we may envision relaxing this condition, especially with respect to the time dimension, so as to have a geometric representation of the time arrow)

4. Strong convexity: the Hessian matrix $\frac{\partial^2}{\partial y^\mu \partial y^\nu}\left(\frac{1}{2}F^2\right)$ is positive-definite at every point of the tangent bundle TM except at the origin $(TM \backslash 0)$.

Intuitively, the value of $F(x, y)$ is interpreted as the length of the vector y tangent at x. More formally, the function F can be used to define the length of a curve indexed by time:

$$s = \int_{t=a}^{t=b} F\left(x(t), y(t)\right) dt \text{ where } y(t) = \frac{dx}{dt}(t). \tag{2.32}$$

as a pseudo-Riemannian metric.

[4] If this assumption is relaxed, then the metric space if called a *Lagrange space*.

If the curve is parametrized in terms of another parameter $\tau = \tau(t)$, for $c = \tau(a) \leq \tau \leq d = \tau(b)$, then the length is given as:

$$s = \int_{\tau=a}^{\tau=b} F(x(\tau), y(\tau))\, d\tau \text{ where } y(\tau) = \frac{dx}{d\tau}(\tau). \tag{2.33}$$

Note that the length integral is independent of the parametrization if and only if condition 2 is valid. Indeed, we have then:

$$s = \int F\left(x, \frac{dx}{dt}\right) dt = \int F\left(x, \frac{dx}{d\tau}\frac{d\tau}{dt}\right) dt = \int F\left(x, \frac{dx}{d\tau}\right)\frac{d\tau}{dt}dt = \int F\left(x, \frac{dx}{d\tau}\right) d\tau. \tag{2.34}$$

We now derive a fundamental proposition that satisfies the metric function.

Proposition 1 *The following relationship holds between the squared value of the metric function and the Hessian of the metric function:*

$$F^2(x,y) = \frac{1}{2} y^\mu y^\nu \frac{\partial^2 F^2(x,y)}{\partial y^\mu \partial y^\nu} \tag{2.35}$$

Proof. By differentiating with respect to p the relation

$$pF\left(x^1, ..., x^\mu, ..., x^n, y^1, ..., y^\mu, ..., y^n\right) = F\left(x^1, ..., x^\mu, ..., x^n, y^1, ..., py^\mu, ..., y^n\right)$$

we obtain:

$$\begin{aligned} F(x,y) &= \frac{\partial F\left(x^1, ..., x^\mu, ..., x^n, y^1, ..., py^\mu, ..., y^n\right)}{\partial p} \\ &= \frac{\partial F\left(x^1, ..., x^\mu, ..., x^n, y^1, ..., py^\mu, ..., y^n\right)}{\partial y^\mu}\frac{\partial\left(py^\mu\right)}{\partial p} = \frac{\partial F(x,y)}{\partial y^\mu} y^\mu \end{aligned} \tag{2.36}$$

which is Euler theorem for homogeneous functions of degree 1. By differentiating the relation $p^2 F(x,y) = F\left(x^1, ..., x^\mu, ...x^\nu, ..., x^n, y^1, ..., py^\mu, .., py^\nu, ..., y^n\right)$ twice with respect to p, we also obtain:

$$\frac{\partial^2 F(x,y)}{\partial y^\mu \partial y^\nu} = 0 \tag{2.37}$$

Now, taking the derivative of the squared metric function with respect to y^μ, we have that:

$$\frac{\partial F^2(x,y)}{\partial y^\mu} = 2F(x,y)\frac{\partial F(x,y)}{\partial y^\mu}$$

Differentiating again with respect to y^ν:

$$\begin{aligned}\frac{\partial^2 F^2(x,y)}{\partial y^\mu \partial y^\nu} &= 2\frac{\partial F(x,y)}{\partial y^\nu}\frac{\partial F(x,y)}{\partial y^\mu} + \underbrace{2F(x,y)\frac{\partial^2 F(x,y)}{\partial y^\mu \partial y^\nu}}_{=0 \text{ from equation 2.37}} \\ &= 2F^2(x,y)\frac{1}{y^\mu y^\nu} \text{ from equation 2.36}\end{aligned}$$

■

This result allows us to define the following covariant tensor of order 2 $f_{\mu\nu}(x,y) = \frac{1}{2}y^\mu y^\nu \times \frac{\partial^2 F(x,y)}{\partial y^\mu \partial y^\nu}$, which will be used to calculate norms of vectors and distances on the manifold M. Indeed, we have $ds \equiv F(x,y) = \sqrt{f_{\mu\nu}(x,y)\, y^\mu y^\nu}$. As explained above, we use in general relativity the framework of standard *Riemannian* geometries, where the metric tensor only depends on coordinates on the spacetime manifold and not on coordinates on the tangent space. In other words, the metric only depends on the position but not on the velocity vector and we write $ds = \sqrt{g_{\mu\nu}(x)\, y^\mu y^\nu}$. The general case where $ds \equiv F(x,y) = \sqrt{f_{\mu\nu}(x,y)\, y^\mu y^\nu}$ defines a broader class of geometries, known as *Finslerian* geometries, named after Paul Finsler, a German and Swiss mathematician who received a PhD in 1918 at the University of Göttingen under the supervision of Constantin Carathéodory, with a focus on extending Riemannian geometry to more general metric specifications [12].

Connections and Covariant Differentiation

Intuitively, the difficulty in performing differential and integral calculus on a curved manifold M is that the tangent vector to a curve on the manifold does not belong to the manifold M itself, but belongs instead to its tangent bundle TM. From this arises the need to formally define how to transport a vector along a curve in a parallel and

consistent manner. An *affine connection* is introduced for transporting tangent vectors to a manifold from one point to another along a curve. An affine connection is typically given in the form of a *covariant derivative*, which defines how to operate the infinitesimal transport of a vector field in a given direction. More formally, let us consider the impact of a change of coordinates $x^\mu \longrightarrow x'^\mu$ on a scalar field ϕ. From the conventional chain rule, we have:

$$\frac{\partial \phi}{\partial x^\mu} \longrightarrow \frac{\partial \phi}{\partial x'^\mu} = \frac{\partial x^\mu}{\partial x'^\mu}\frac{\partial \phi}{\partial x^\mu}, \tag{2.38}$$

or equivalently:

$$\partial_\mu \phi \longrightarrow \partial'_\mu \phi = \frac{\partial x^\mu}{\partial x'^\mu}\partial_\mu \phi, \tag{2.39}$$

where we have used the shorthand notation $\partial_\mu \phi = \frac{\partial \phi}{\partial x^\mu}$ for the partial derivative of ϕ with respect to x^μ and $\partial'_\mu \phi = \frac{\partial \phi}{\partial x'^\mu}$ for the partial derivative of ϕ with respect to x'^μ. If we now apply the change of coordinate $x^\mu \longrightarrow x'^\mu$ to a vector field V^ν, we obtain:

$$\partial_\mu V^\nu \longrightarrow \partial'_\mu V^{\nu'} = \left(\frac{\partial x^\mu}{\partial x'^\mu}\partial_\mu\right)\left(\frac{\partial x'^\nu}{\partial x^\nu}V^\nu\right) \tag{2.40}$$

$$= \frac{\partial x^\mu}{\partial x'^\mu}\frac{\partial x'^\nu}{\partial x^\nu}\partial_\mu V^\nu + \underbrace{\frac{\partial x^\mu}{\partial x'^\mu}\frac{\partial^2 x'^\nu}{\partial x^\mu \partial x^\nu}V^\mu}_{\text{non-tensorial terms}}, \tag{2.41}$$

We would like to define a new form of derivative operator ∇_μ, known as the *covariant derivative* operator, which would obey the tensorial transformation law:

$$\nabla_\mu V^\nu \longrightarrow \nabla_{\mu'} V'^\nu = \frac{\partial x^\mu}{\partial x'^\mu}\frac{\partial x^\nu}{\partial x'^\nu}\nabla_\mu V^\nu. \tag{2.42}$$

To do so, we introduce the so-called *Christoffel symbols* $\Gamma^\nu_{\mu\lambda}$, or *connection coefficients*, which are required ingredients in the definition of the covariant derivative as the usual partial derivative plus a linear correction:

$$\nabla_\mu V^\nu = \partial_\mu V^\nu + \Gamma^\nu_{\mu\lambda} V^\lambda.$$

So when given a particular metric $g_{\mu\nu}$, the first step consists of calculating the connection coefficients so that we can take covariant derivatives. It can be shown that to cancel out the non tensorial terms defined in equation (2.41), the connection coefficients $\Gamma^{\nu}_{\mu\lambda}$ need to obey the following transformation law under a coordinate transformation $x^{\mu} \rightarrow x'^{\mu}$ (from which we see that these connection coefficients $\Gamma^{\nu}_{\mu\lambda}$ are not tensors):

$$\Gamma^{\nu'}_{\mu'\lambda'} = \frac{\partial x^{\mu}}{\partial x'^{\mu}} \frac{\partial x^{\lambda}}{\partial x'^{\lambda}} \frac{\partial x'^{\nu}}{\partial x^{\nu}} \Gamma^{\nu}_{\mu\lambda} - \frac{\partial x^{\mu}}{\partial x'^{\mu}} \frac{\partial x^{\lambda}}{\partial x'^{\lambda}} \frac{\partial^2 x'^{\nu}}{\partial x^{\mu} \partial x^{\lambda}} \tag{2.43}$$

The connection coefficients are not uniquely defined, but they can be chosen to admit a natural expression defined from the metric $g_{\mu\nu}$ and its derivatives:

$$\Gamma^{\sigma}_{\mu\nu} = \frac{1}{2} g^{\sigma\rho} \left(\partial_{\mu} g_{\nu\rho} + \partial_{\nu} g_{\rho\mu} - \partial_{\rho} g_{\mu\nu} \right). \tag{2.44}$$

in which case they are known as the *Christoffel symbols*. The choice in equation (2.44) implies that the covariant derivative of the metric and its inverse are always zero, a property known as *metric compatibility*:

$$\nabla_{\rho} g^{\mu\nu} = 0. \tag{2.45}$$

Parallel Transport, Geodesics and Curvature

The main insight in General Relativity is that spacetime is a curved manifold, and that it is the mass/energy content of the universe that defines its local curvature. Intuitively, curvature is the amount by which a manifold deviates from being a flat geometry. A key distinction exists between an *extrinsic* definition of curvature, which is defined by embedding the manifold within another higher-dimensional space, and an *intrinsic* definition of curvature, which is measured at each point of the manifold from the properties of the manifold itself, without the need to resort to a higher-dimensional space. In Riemannian geometry, the intrinsic approach to the definition of curvature is related to the concept of *parallel transport* of a vector. Intuitively, if a vector is left unchanged after being parallel-

transported on a closed infinitesimal loop around a certain point of the manifold, then the manifold is flat at this point. On the other hand, if the action of parallel transport along the closed infinitesimal loop alters the vector, then the manifold is curved, and the curvature can be related to the degree of alteration of the vector under the action of parallel transport. Formally, we define parallel transport of the tensor T along the path $x^{\mu}(\lambda)$ to be the requirement that the covariant derivative of T along the path be zero:

$$\frac{dx^{\sigma}}{d\lambda}\nabla_{\sigma}T^{\mu_1\mu_2\cdots\mu_k}_{\nu_1\nu_2\ldots\nu_l} = 0. \tag{2.46}$$

For a vector V, the equation of parallel transport takes the form:

$$\frac{d}{d\lambda}V^{\mu} + \Gamma^{\mu}_{\sigma\rho}\frac{dx^{\sigma}}{d\lambda}V^{\rho} = 0, \tag{2.47}$$

from which we see that the notion of parallel transport is dependent upon the choice of the connection. Before discussing how parallel transport can be used to propose an intrinsic definition of curvature, we first remark that parallel transport can also allow us to define *geodesics*. This is an important concept since we will see below that Einstein field equation implies that freely falling test particles follow geodesics. Intuitively, a geodesic is the closest approximation in a curved manifold of a straight line in a flat manifold. Formally a geodesic is a parametrized curve $x^{\mu}(\lambda)$ that minimizes the distance between two points A and B: $L = \int_A^B ds = \int_A^B \sqrt{g_{\mu\nu}\frac{dx^{\mu}(\lambda)}{d\lambda}\frac{dx^{\nu}(\lambda)}{d\lambda}}d\lambda$. It turns out that this definition is equivalent to another definition of a geodesic if (and only if) the connection used is the Christoffel connection (see for example [39]). This second definition is expressed as follows: a geodesic is a curve along which the tangent vector is parallel transported, which generalizes the obvious result that a straight line in a flat space is a path that parallel-transports its own tangent vector. Since the tangent vector to a path $x^{\mu}(\lambda)$ is $\frac{dx^{\mu}(\lambda)}{d\lambda}$ we obtain from 2.47 that the Euler-Lagrange geodesic equation solution to the variational

problem

$$Min \int_A^B \sqrt{g_{\mu\nu} \frac{dx^\mu(\lambda)}{d\lambda} \frac{dx^\nu(\lambda)}{d\lambda}} d\lambda \tag{2.48}$$

can be written as:

$$\frac{d^2x^\mu}{d\lambda^2} + \Gamma^\mu_{\nu\sigma} \frac{dx^\nu}{d\lambda} \frac{dx^\sigma}{d\lambda} = 0. \tag{2.49}$$

if λ is an affine parameter.

We are now ready to introduce an intrinsic definition of curvature. To do this, we consider the parallel transport of a vector V along a closed infinitesimal loop defined in terms of two adjacent vectors A and B. Thus the vector V is first parallel-transported in the direction of A, then in the direction of B, and then backward along the directions of A and B so as to return to the starting point. This allows us to introduce the *Riemann curvature tensor* $R^\rho_{\sigma\mu\nu}$ through the following relationship:

$$\delta V^\rho = R^\rho_{\sigma\mu\nu} V^\sigma A^\mu B^\nu, \tag{2.50}$$

where δV is the change in the vector V occurred from parallel-transporting the vector along the infinitesimal closed loop defined by the two vectors A and B. Using the characterization of parallel transport equation 2.47, it is possible to obtain an explicit expression for the Riemann curvature tensor as a function of the connection coefficients:

$$R^\sigma_{\mu\alpha\beta} = \partial_\alpha \Gamma^\sigma_{\mu\beta} - \partial_\beta \Gamma^\sigma_{\mu\alpha} + \Gamma^\sigma_{\alpha\lambda} \Gamma^\lambda_{\mu\beta} - \Gamma^\sigma_{\beta\lambda} \Gamma^\lambda_{\mu\alpha}, \tag{2.51}$$

from which we can see that $R^\sigma_{\mu\alpha\beta}$ is antisymmetric in its two last indices

$$R^\sigma_{\mu\alpha\beta} = -R^\sigma_{\mu\beta\alpha}.$$

There are two contractions of the Riemann curvature tensor that are extremely useful,

namely the Ricci tensor $R_{\alpha\beta}$ and the Ricci scalar R:

$$R_{\alpha\beta} \equiv R^{\lambda}_{\alpha\lambda\beta} \tag{2.52}$$

$$R \equiv R^{\lambda}_{\lambda} = g^{\mu\nu} R_{\mu\nu} \tag{2.53}$$

It can also be shown that the Riemann tensor satisfies a differential identity known as the Bianchi identity:

$$\nabla_{\lambda} R_{\mu\nu\rho\sigma} = 0 \tag{2.54}$$

If we now define the Einstein tensor as:

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}, \tag{2.55}$$

then the Bianchi identity implies that:

$$\nabla^{\mu} G_{\mu\nu} = 0. \tag{2.56}$$

Einstein Equation

The fundamental equation of motion in general relativity is the Einstein equation:

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \tag{2.57}$$

where the constant $\frac{8\pi G}{c^4}$ allows one to recover the standard gravitational potential in the Newtonian limit of slow motion and weak gravitational field, and where $T_{\mu\nu}$ is the mass/energy tensor, which describes the mass/energy content of the universe. For example, the stress-energy tensor for a perfect fluid admits the following general form in the rest frame of the fluid:

$$T_{\mu\nu} = (\rho + p)\, U_{\mu} U_{\nu} + p g_{\mu\nu}, \tag{2.58}$$

where ρ and p are the rest-frame energy and momentum densities of the fluid and U_μ is the fluid 4-velocity defined in 2.23

$$U_\mu = \frac{dx^\mu}{d\tau} \tag{2.59}$$

and where τ is the fluid proper time, that is the time measured by an observer in the fluid rest frame. The Einstein equation 2.57 is a gravitational field equation. As John Wheeler colorfully put it, "matters tells spacetime how to curve and spacetime tells matter how to move". In this description, gravitation is no longer regarded as a force, but as a geometric manifestation of the curvature of spacetime. Taking the trace of the Einstein equation, we obtain:

$$-R = \frac{8\pi G}{c^4} T \tag{2.60}$$

so that Einstein equation can be rewritten as:

$$R_{\mu\nu} = \frac{8\pi G}{c^4}\left(T_{\mu\nu} - \frac{1}{2} T g_{\mu\nu}\right). \tag{2.61}$$

An important specific situation is the vacuum situation, where $T_{\mu\nu} = 0$ so that Einstein equation simply becomes:

$$R_{\mu\nu} = 0. \tag{2.62}$$

It has been shown by David Hilbert [94] that the Einstein gravitational field equation 2.57 can be derived through the principle of least action, which allows for an unification of general relativity with other classical field theories such as Maxwell theory. To see this, we define the Hilbert–Einstein action as:

$$S = \frac{1}{16\pi G c^{-4}} \int R\sqrt{-g} d^4x, \tag{2.63}$$

an expression which can be extended to account for the presence of a matter field.

A more general version of Einstein equation is a version involving the so-called cosmological constant Λ:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu} \tag{2.64}$$

As is well-known, the cosmological constant had originally been introduced by Albert Einstein [74] to allow for a static solution of the gravitational field equation when applied to the universe assumed to be filled with dust. Albert Einstein eventually called the introduction of this term its "biggest blunder" after Edwin Hubble's 1929 discovery [97] that all galaxies (outside the Local Group) exhibited redshift and hence had recession velocities, which suggests that they were moving away from the Milky Way and from each other, implying an overall expanding universe.

In fact, the left-hand side of 2.64 is the most general local, coordinate-invariant, divergenceless, symmetric, two-index tensor that can be constructed solely from the metric and its first and second derivatives. As a result, the cosmological constant should be regarded as a legitimate addition to the gravitational field equations, and as a parameter to be constrained by observation. From the physical standpoint, there is an equivalence between GR with cosmological constant and GR with vacuum energy in addition to matter (and radiation). To see this, we split the energy-momentum tensor into a term describing matter/energy and a term describing the vacuum:

$$T_{\mu\nu} = T^{mat}_{\mu\nu} + T^{vac}_{\mu\nu}. \tag{2.65}$$

In order to maintain Lorentz invariance, vacuum energy should also have no preferred direction. Therefore the first term in the perfect fluid energy tensor must be zero, requiring $p_{vac} = -\rho_{vac}$. As a result, we obtain from equation 2.58:

$$T_{\mu\nu} = -\rho_{vac}g_{\mu\nu}, \tag{2.66}$$

and Einstein's equation including vacuum energy (without the cosmological constant)

becomes:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \frac{8\pi G}{c^4}\left(T_{\mu\nu}^{\text{matter}} - \rho_{vac}g_{\mu\nu}\right). \tag{2.67}$$

By comparison with equation 2.64, we confirm that the cosmological constant can be identified with vacuum energy as long as:

$$\rho_{vac} = \Lambda\frac{c^4}{8\pi G}, \tag{2.68}$$

a quantity which is sometimes also denoted byρ_Λ. In the next section, we shall discus the relevance of the introduction of the cosmological constant in cosmology, in particular for accounting for the empirical observation that the expansion of the universe is accelerating.

2.1.2 Application of GR to Cosmology

In this section, we provide an introduction to the standard ΛCDM (lambda cold dark matter) cosmological model, a parametrization of the Big Bang cosmological model in which the universe contains a cosmological constant, denoted by Lambda (Λ), associated with dark energy, and cold dark matter. This simple model provides a reasonably good description of the main observed features of the universe, and is obtained by applying general relativity as the assumed correct theory of gravity up to the cosmological scales. We also review how the ΛCDM model can be extended by adding cosmological inflation so as to account for a number of otherwise unexplained features of the observable universe. This introduction to cosmology will prove useful in defining some basic quantities involved in the analysis of gravitational wave signals (such as the Hubble constant, the critical density needed to close the universe, etc.). It will also prove useful in our analysis of the sources of GWs of cosmological origin (such as phase transitions, cosmic strings, etc.). Classical references on these subjects are [174] and [135].

Friedmann-Lemaître-Robertson-Walker Metric and Friedmann Equations

Cosmology is the analysis of the largest-scale structures and dynamics of the Universe. Cosmology as a science originated with the *Copernican principle*, which is a working assumption that can be regarded as a modified cosmological extension of Copernicus' heliocentric universe. Under this modified Copernican principle neither the Sun nor the Earth are in a central, specially favored position in the universe. This modified *Copernican principle* can be related to two important properties that spacetime regarded as a manifold is assumed to enjoy, namely *isotropy* and *homogeneity*. Roughly speaking, a manifold is homogeneous if it is invariant under any translation along a coordinate, and it is isotropic if it is invariant under any rotation of a coordinate into another coordinate. Intuitively, isotropy is a local concept stating that a manifold is isotropic around a given point if the geometry on the manifold is the same regardless of direction as seen from this point (see for example [39] for a formal definition). Although the universe is clearly not homogeneous at smaller scales, there is ample observational evidence that the universe is highly homogeneous, and isotropic as regarded from the Earth, at a sufficiently large scale (e.g., at a scale larger than 250 million light years). In particular, a statistical analysis of the cosmic microwave background (CMB), the thermal radiation left over from the time of photon decoupling in Big Bang cosmology (see more details below), suggests that this radiation is isotropic to roughly one part in 100,000 and the root mean square variations around the mean value of 2.72548 K are only 18×10^{-6} K, after subtracting out a dipole anisotropy from the Doppler shift of the background radiation caused by the peculiar velocity of the Earth relative to the comoving cosmic rest frame (see for example [179]). Isotropy and homogeneity are two distinct properties, which can be related as follows. If spacetime is isotropic everywhere, then it is homogenous. Conversely, if it is isotropic around a certain point and also homogenous, then it must be isotropic around every point. Because the universe appears highly isotropic from the Earth, an application of the Copernician principle implies that the universe is assumed to be isotropic and

therefore homogeneous. Since we are at no special place in the universe, observers in other places of the universe should also observe isotropy.

Homogeneity will be used to model the matter content of the universe as a *perfect fluid* with no shear stresses, viscosity, or heat conduction, and therefore fully characterized by two parameters only, its mass density ρ and pressure p. The stress-energy tensor for the perfect fluid admits the expression in equation 2.58. Note that the trace of the stress-energy tensor is:

$$T = T^{\mu}_{\mu} = -\rho + 3p. \tag{2.69}$$

When taken together, the two assumptions that the universe is homogeneous and isotropic are known as the *Cosmological Principle*, which is the main underlying principle used when applying GR to the analysis of the universe as a whole. In particular, the cosmological principal can be used to provide constraints on the form of spacetime geometry. More precisely, if we assume that the universe is spatially homogenous and isotropic but is allowed to evolve in time, it can be shown (see [174] for formal arguments) that spacetime geometric is captured by a family of metric functions, known as the Friedmann-Lemaître-Robertson-Walker metric functions, which can be written as follows in spherical coordinates (r, θ, ϕ):

$$ds^2 = -c^2 dt^2 + a^2(t)\left[f(r)\, dr^2 + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right], \tag{2.70}$$

where

$$f(r) = \left(1 - kr^2\right)^{-1}. \tag{2.71}$$

Here k is the value of curvature, assumed constant to ensure homogeneity, and r can always be renormalized to ensure that

$$k \in \{-1, 0, 1\}. \tag{2.72}$$

The case $k = -1$ corresponds to an open universe with a constant negative curvature;

the case $k = 0$ correspond to a flat universe with zero curvature; and the case $k = +1$ corresponds to a closed universe with positive curvature. Out of 16 coefficients $g_{\mu\nu}$ there are 10 independent coefficients due to symmetry. In the Robertson-Walker metric, only 4 out of the 10 coefficients are non-vanishing :

$$g_{tt} = -c^2 \tag{2.73}$$

$$g_{rr} = \frac{a^2(t)}{1-kr^2} \tag{2.74}$$

$$g_{\theta\theta} = a^2(t)\, r^2 \tag{2.75}$$

$$g_{\phi\phi} = a^2(t)\, r^2 \sin^2\theta \tag{2.76}$$

From the metric, we can compute the Christoffel symbols and subsequently the components of the Riemann tensor, from which can be obtained the Ricci tensor as well as the Ricci scalar, which are the needed ingredients for the left-hand side of Einstein's field equation. Using the notation $\dot{a} \equiv \frac{da}{dt}$, we obtain the following non-zero components of the Ricci tensor:

$$R_{00} = -\frac{3\ddot{a}}{ac^2} \tag{2.77}$$

$$R_{ii} = -\frac{2k + \frac{a\ddot{a}}{c^2} + \frac{2\dot{a}^2}{c^2}}{a^2} g_{ii} \tag{2.78}$$

and the Ricci scalar is:

$$R = \frac{6}{c^2 a^2}\left[a\ddot{a} + \dot{a}^2 + kc^2\right] \tag{2.79}$$

We may now apply the Einstein equation with the cosmological constant:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}, \tag{2.80}$$

and we obtain the so-called *Friedmann equation* from the $\mu\nu = ii$ components (see for example [39]):

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3}. \tag{2.81}$$

We also obtain for the $\mu\nu = 00$ component:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3}, \tag{2.82}$$

which is sometimes called the *Friedmann acceleration equation,* or the *second Friedmann equation.*

It is also useful to consider the zero component of the conservation of energy equation:

$$\nabla_\mu T_0^\mu = 0, \tag{2.83}$$

which yields:

$$\partial_\mu T_0^\mu + \Gamma^\mu_{\mu\lambda} T_0^\lambda - \Gamma^\lambda_{\mu 0} T_\lambda^\mu = 0, \tag{2.84}$$

or

$$\dot{\rho}c^2 + 3\frac{\dot{a}}{a}(\rho c^2 + p) = 0 \Leftrightarrow \frac{\dot{\rho}}{\rho} = -3\,(1+w)\,\frac{\dot{a}}{a}, \tag{2.85}$$

where $w = p/\rho c^2$. If w is a constant, the conservation of energy equation 2.85 can be integrated to give:

$$\rho \propto a^{-3(1+w)}. \tag{2.86}$$

The two Friedmann equations 2.81 and 2.82 involve three unknowns, namely a, p and ρ. A third equation is therefore needed, which in the case of a perfect fluid can be obtained by *the equation of state* 2.86 which relates p and ρ. We make a distinction between three cases of interest for cosmology.

- The first case corresponds to a universe driven by matter, which is defined as any set of collisionless particles for which we can assume a zero pressure $p = 0$. In this *matter-dominated* case, the conservation of energy equation 2.86 implies that the density matter in the universe $\rho_M \propto a^{-3}$ as the universe expands.

- The second case corresponds to an earlier epoch, when the universe was much hotter and denser implying the presence of strong interactions so that pressure does not

vanish and can be assumed to be related to the density parameter by $p = \frac{1}{3}\rho c^2$. In this *radiation-dominated* case, the conservation of energy equation 2.86 implies that the density of radiation in the universe $\rho_R \propto a^{-4}$ as the universe expands.

- The third case of interest is the situation where we assume that $p = -\rho c^2$. In this vacuum-dominated case, which corresponds to the present epoch, the conservation of energy equation 2.86 implies that the energy density is constant (proportional to a^0) as the universe expands.

There are a number of quantities of interest for cosmology that can be defined at this stage. We first introduce the rate of expansion of the universe, which is known as the Hubble parameter:

$$H = \frac{\dot{a}}{a} \tag{2.87}$$

The value of this parameter at the current time is known as the Hubble constant, denoted by H_0, which is estimated at 70 ± 10 km/sec/Mpc, where a megaparsec Mpc $= 3.09 \times 19^{24}$ cm. Given the uncertainty around this parameter, it is sometimes written as:

$$H_0 = 100h \text{ km/sec/Mpc}, \tag{2.88}$$

where h is taken to be 70% as a base case estimate. Related to the Hubble constant are the *Hubble time*:

$$t_H \equiv \frac{1}{H_0} \simeq 4.4 \times 10^{17} \text{ s}, \tag{2.89}$$

and the *Hubble length*:

$$l_H \equiv \frac{c}{H_0} \simeq 1.3 \times 10^{26} \text{ s}. \tag{2.90}$$

From an observational standpoint, the scale factor $a(t)$ can be related to the wavelength shift of a light signal. Let us define the redshift as:

$$z = \frac{\Delta\lambda}{\lambda} = \frac{\lambda_{now} - \lambda_{then}}{\lambda_{then}}, \tag{2.91}$$

or

$$1 + z = \frac{\lambda_{now}}{\lambda_{then}},$$

where λ_{now} is the wavelength of an electromagnetic signal measured by an observer at the current time and λ_{then} is the wavelength of the same signal when it has been emitted. Using the geodesic equation for a light wave, simple arguments show that

$$\frac{\lambda_{now}}{\lambda_{then}} = \frac{a_{now}}{a_{then}}. \tag{2.92}$$

Given that the current time is t_0 with $a(t_0) = 1$, we obtain the basic relation:

$$1 + z = \frac{1}{a_{then}}. \tag{2.93}$$

Thus, at the redshift $z = 1$, the universe had a size half as the present one. As another example, the "photon decoupling time" when the universe became transparent to light (see next section) is said to occur at $z = 1100$. The scale factor a can also be used to define the *deceleration parameter* q, a dimensionless measure of the cosmic acceleration of the expansion of space in a Friedmann-Lemaître-Robertson-Walker universe, which is given by:

$$q = -\frac{a\ddot{a}}{\dot{a}^2}. \tag{2.94}$$

The acceleration of expansion $\ddot{a}$ can be computed as follows. Let us start from the Friedmann equation 2.81 without the cosmological constant ($\Lambda = 0$) in a flat universe ($k = 0$), which can be written as:

$$\dot{a} = a\sqrt{\frac{8\pi G}{3}\rho}. \tag{2.95}$$

Taking the time derivative of this equation and replacing $\dot{\rho}$ according to the energy

conservation equation 2.85, we obtain:

$$\ddot{a} = -\sqrt{\frac{8\pi G}{3}}\frac{\dot{a}}{2\sqrt{\rho}}\left(\rho + 3p\right). \tag{2.96}$$

Note that the derivative of the Hubble parameter can be written in terms of the deceleration parameter:

$$\frac{\dot{H}}{H^2} = -\left(1 + q\right). \tag{2.97}$$

Another useful quantity related to the Hubble parameter is the critical density:

$$\rho_{crit} \equiv \frac{3H^2}{8\pi G}, \tag{2.98}$$

which is defined as the value of the density parameter required to have $k = 0$. This can be seen from the first Friedmann equation without the cosmological constant ($\Lambda = 0$), which we can rewrite as:

$$\frac{kc^2}{\dot{a}^2} = \frac{8\pi G}{3H^2}\rho - 1. \tag{2.99}$$

We also may define the *critical density parameter* as:

$$\Omega = \frac{\rho}{\rho_{crit}} \equiv \frac{8\pi G}{3H^2}\rho. \tag{2.100}$$

The sign of k and the geometry of the universe is therefore defined as a function of the density parameter:

$$\rho < \rho_{crit} \Leftrightarrow \Omega < 1 \Leftrightarrow k < 0 \Leftrightarrow \text{open universe} \tag{2.101}$$

$$\rho = \rho_{crit} \Leftrightarrow \Omega = 1 \Leftrightarrow k = 0 \Leftrightarrow \text{flat universe} \tag{2.102}$$

$$\rho > \rho_{crit} \Leftrightarrow \Omega > 1 \Leftrightarrow k > 0 \Leftrightarrow \text{closed universe} \tag{2.103}$$

The density parameter is often divided into components so as to provide a quantitative measure of the relative contribution of various sources of matter/energy to the energy

content of the universe. If we introduce:

$$\Omega_M = \frac{\rho_M}{\rho_{crit}}, \tag{2.104}$$

$$\Omega_R = \frac{\rho_R}{\rho_{crit}}, \tag{2.105}$$

$$\Omega_\Lambda = \frac{\rho_\Lambda}{\rho_{crit}}, \tag{2.106}$$

where ρ_M is the density of matter, ρ_R is the density of radiation, and ρ_Λ is the density for vaccum, the first Friedmann equation 2.81 gives:

$$\frac{H^2}{H_0^2} = \Omega_M a^{-3} + \Omega_R a^{-4} + \Omega_\Lambda, \tag{2.107}$$

which provides a useful measure of the evolution of the Hubble parameter as a function of the evolution of the density for each type of matter/energy as the universe expands. For Ω_R it is often useful to express the quantity as a function of the frequency of the source of radiation. Given that a logarithmic scale is typically used, we set:

$$\Omega_R = \int \Omega_R(f)\, d\ln f, \tag{2.108}$$

with:

$$\Omega_R(f) \equiv \frac{1}{\rho_{crit}} \frac{d\rho_R}{d\ln f}, \tag{2.109}$$

a quantity which will be used in chapter 3 in the analysis of gravitational wave signals.

The ΛCDM Cosmological Model with Cosmological Inflation

Historically, the first popular cosmological model has been the standard Hot Big Bang model (without dark matter and dark energy), which uses the FLRW metric 2.70, the Friedmann equations 2.81 and 2.82, as well as the cosmological equation of state 2.86 to describe the observable universe from an origin known as the Big Bang, a hot and dense initial state starting from which the universe has been expanding over time to present.

One of the key successful predictions of this cosmological model is the existence of a *cosmic microwave background* (CMB), defined as the thermal radiation left over from the time of recombination in Big Bang cosmology. As the universe expanded, both the plasma and the radiation filling it grew cooler until a point, estimated to be at 10^{13} seconds (or about $400,000$ years) after the Big Bang, known as the *recombination epoch*, when the universe cooled enough for protons and electrons to start combining and forming neutral hydrogen atoms. Since these atoms could no longer absorb the thermal radiation, the universe became transparent as photons started to travel freely through space rather than constantly being scattered by electrons and protons in plasma. This photon decoupling effect is the source of the relic radiation. The photons that existed at that time have been propagating ever since while growing fainter and less energetic given that the expansion of space causes their wavelength to increase over time. The accidental discovery of a cosmic microwave background with a thermal black body spectrum at a temperature of $2.72548 \pm 0.00057\ K$ in 1964 by American radio astronomers Arno Penzias and Robert Wilson [136] came as a confirmation of this key prediction of the Big Bang cosmology, which then became the commonly accepted cosmological model. Despite its success in accounting for a large body of cosmological observations, the standard Big Bang model fails, however, to explain a number of puzzles or problems. The two most important of these problems are known as the *horizon problem* and the *flatness problem*, which we briefly summarize below. Other problems exist, such as the *magnetic monopole problem* or the *entropy problem*, for which cosmic inflation can also possibly provide a reasonable explanation [41].

- The horizon problem (sometimes known as the homogeneity problem or the causality problem) is the problem of determining why the universe appears statistically homogeneous and isotropic [128], which justifies the application of the cosmological principle. As recalled above, the cosmic microwave background has been found isotropic to roughly one part in 100,000 and the root mean square variations around the mean value of $2.72548\ K$ are only $18 \times 10^{-6}\ K$ [179]. In a standard Big Bang

model, the whole surface of last scattering is not in causal contact and therefore widely separated regions of the observable universe cannot have equilibrated given that the transfer of information (or energy, heat, etc.) required to ensure homogeneity can occur, at most, at the speed of light.

- The flatness problem (sometime also called one of the Dicke coincidences, along with the cosmological constant problem [60]) refers to the problem of obtaining measures for the density of the universe that are extremely close to the critical density needed to close the universe, which in turns requires an extreme fine tuning of the constant Ω at earlier times. From the observation of the anisotropies, or angular scale of fluctuations, of the Cosmic Microwave Background (CMB) radiation, Ω has indeed been found to be equal to 1 ± 0.01 [162], a result which has been independently confirmed by Type-Ia supernovae surveys and the analysis of their redshift at different distances from Earth [127]. That Ω is so close to the critical value of 1 today is extremely surprising when the situation is assessed at earlier times. To see this, note that equation 2.99 implies:

$$k \propto a^2 \rho \left(1 - \Omega^{-1}\right) \sim a^2 \rho \left(\Omega - 1\right). \tag{2.110}$$

Now, $a^2\rho$ is not a constant since radiation density ρ_R falls of as a^{-4} as the universe expands, while matter density ρ_M falls of as a^{-3}. From the time of the Big Bang to the present time, $a^2\rho$ is therefore expected to have decreased by approximately 60 orders of magnitude in a standard FLRW cosmology. This in turn implies that the constant k would have increased by approximately 60 orders of magnitude, and therefore

$$\left|\Omega\left(t = t_0\right) - 1\right| \leq 10^{-2} \Rightarrow \left|\Omega\left(t = t_{\mathrm{BB}}\right) - 1\right| \leq 10^{-62} \tag{2.111}$$

where t_{BB} is some time just after the Big Bang. This is an extreme example of a *fine tuning* problem, which raises the question of how to explain that the energy density has taken a value so extremely close to the critical value required to close

the universe. In fact any small deviation of Ω from 1 in the early universe would have been magnified during billions of years of expansion to create a current density very far from the critical value.

In a seminal paper [90], Alan Guth has proposed an extension of the standard cosmological model as a possible solution to the horizon and flatness problems. The model, known as *cosmological inflation model* or *inflation model* in short, proposes that within 10^{-29} seconds after the Big Bang occurred an exponential expansion of space with a scale factor a increasing exponentially by at least 26 orders of magnitude. This comic inflation scenario would solve the horizon problem since it implies that the entire observable universe today would have started as a small causally connected region. It would also solve the flatness problem since the exponential increase in the scale factor would have led to an exponential decrease in $\Omega - 1$. Hence, under this scenario, the density today would be extremely close to 1 whatever was the initial density value. Since the flatness and horizon problems can be explained by an early stage of accelerated expansion, it is useful to analyze the physical conditions under which this early accelerated expansion can occur. This can be done from equation 2.96, which states that the universe is in accelerating expansion if $(\rho + 3p) < 0$. We conclude that during inflation the pressure should be negative and smaller than $-\rho/3$. A cosmological constant would satisfy this property since $p = -\rho$. During a fully Λ-dominated stage, however, the energy density stays constant and never decays, implying that inflation will be indefinite. The most straightforward possibility for inflation to eventually come to an end is to consider a scalar field (called the inflaton) slowly rolling down in a very flat valley. At all times in this slow roll-down process the field can be regarded as a false vacuum state sharing the key properties of the true vacuum state, including that the energy of the field is diluted very slowly with a pressure very close to $-\rho$. In the inflationary stage, the potential energy of the slow-rolling scalar field (the inflaton) starts to dominate the total energy density in the universe, until the quantum fluctuations of the scalar field and of the metric undergo a semi-classical transition. At the end of inflation, the scalar field decays

into particles in a stage called reheating, during which the scalar field decays. After the decay of the scalar field, the universe is dominated by the energy of relativistic particles produced during reheating and enters into the aforementioned radiation-dominated stage. A detailed description of cosmological inflation is beyond the scope of the present discussion (see for example [115] for a review). Here we mostly emphasize that cosmic inflation is expected to have important implications for the generation of gravitational waves. Indeed in its lowest energy state, the quantum inflation field has only the zero point energy. As space expands, the field expands as well and quanta of the gravitational field, called gravitons, are created, a process which is expected to lead to the generation of gravitational waves of extremely low frequency (in the range of 10^{-18}-10^{-12} Hz) as the wavelength of these gravitons expands in the inflating universe. In section 3.1.2, which is dedicated to the stochastic gravitational wave background from cosmological origin, we revisit the question of the primordial stochastic background generated by inflation in more details, and we also discuss the generation of gravitational waves from cosmic strings.

In addition to the puzzles related to the origin of the universe, pure-baryonic models also face serious challenges explaining a variety of observational features of the current universe, such as the flatness of the galactic rotation curves and the mass discrepancy in clusters of galaxies ([135]), which were taken as evidence of the presence of *cold dark matter* in addition to baryonic matter. As a consequence, most research during the 1980s focused on CDM cosmological models involving cold dark matter with critical density in matter around 95% CDM and 5% baryons, which have been found successful at explaining the formation of galaxies and clusters of galaxies (see [31] for a review). A number of problems remained, however, including the fact that the model required a Hubble constant lower than what suggested by observations. These difficulties sharpened with the improved estimates for the CMB anisotropy by COBE in 1992, and several modified CDM models, including the ΛCDM model, came under active consideration through the mid-1990s. [135] is a classic reference for an overview of the historical stages

of development of the standard model (see also [175] for a more recent reference).

The currently accepted ΛCDM model is a version of the standard Big Bang cosmology that contains the cosmological constant, as presented in equations 2.81 and 2.82. The introduction of the cosmological constant was confirmed by the finding from two separate teams of astronomers observing distant type 1a supernovas, one led by the American Saul Perlmutter and the other by the Australians Nick Suntzeff and Brian Schmidt ([137] and [147]), that the deceleration rate of expansion was in fact negative, namely that the universe expansion was accelerating. In addition to explaining earlier observations, the model has made a number of successful predictions including the existence of acoustic oscillations in the density of the visible baryonic matter in the predicted location [85] and the statistical features of weak gravitational lensing, first observed in 2000 by several teams (see [110] for a recent review of observational results regarding cosmic shear, the distortion of images of distant galaxies due to weak gravitational lensing by the large-scale structure in the Universe, and their implications to cosmology). The polarization of the CMB, discovered in 2002 by the Degree Angular Scale Interferometer (DASI) [112], is also a success for the ΛCDM model: in the 2015 Planck data release, there are seven observed peaks in the temperature power spectrum, six peaks in the temperature-polarization cross spectrum, and five peaks in the polarization spectrum, and they all agree with the predictions of the ΛCDM model [46]. Beside, the standard ΛCDM model was found to fit the Wilkinson Microwave Anisotropy Probe (WMAP) temperature and polarization data [162].

2.1.3 Gravitational Waves

Special relativity implies that all interactions, including gravitational interactions, can travel at most at the speed of light and cannot propagate instantly as suggested in Newton's theory. General relativity re-interprets gravitation not as a force, but as a manifestation of spacetime curvature. Taken together these theories imply the existence of some form of "gravitational radiation", known as gravitational waves (GWs), which

are space-time perturbations travelling at the speed of light created by time variations of the *mass quadrupole moment* of the source (see equation 2.178), and which existence has been predicted by Albert Einstein in 1916 [73].

Gravitational Waves in Weak Gravitational Fields

The starting point for the description of GWs is to consider a spacetime endowed with a metric that is deviating very slightly from the flat Minkowski metric, and which can be written as the Minkowski metric of special relativity $\eta_{\mu\nu}$ (equations 2.5-2.7) plus a small metric perturbation $h_{\alpha\beta}$:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{2.112}$$

with:

$$|h_{\mu\nu}| \ll 1. \tag{2.113}$$

In what follows, all indices are raised and lowered with the flat Minkowski metric $\eta_{\mu\nu}$. This metric is consistent with the presence of a weak gravitational field, which would be generated by non compact objects such as non massive stars, or by a compact object located at large distances. The inverse of the Minkowski matrix is the Minkowski matrix itself:

$$\eta^{-1} \times \eta = Id_4, \tag{2.114}$$

which can equivalently be written as:

$$\eta^{\mu\alpha}\eta_{\alpha\nu} = \delta^{\mu}_{\nu}. \tag{2.115}$$

As a result, we have that the inverse matrix of $g_{\mu\nu}$, $g^{\mu\nu}$, can be written as $\eta^{\mu\nu}$ plus a small perturbation $k^{\mu\nu}$. It can easily be shown that $k^{\mu\nu} = -h^{\mu\nu}$ so that we finally have:

$$g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu}, \tag{2.116}$$

where:

$$h^{\mu\nu} \equiv \eta^{\mu\alpha}\eta^{\nu\beta}h_{\alpha\beta}. \tag{2.117}$$

Linearized Form of Einstein Equation We now can start looking at the linearized form of the Einstein equation. For this, we first consider the components of the affine connection (the Christoffel symbols) $\Gamma^{\rho}_{\mu\nu}$ associated with the metric g:

$$\Gamma^{\rho}_{\mu\nu} = \frac{1}{2}\left(\eta^{\rho\lambda} - h^{\rho\lambda}\right)\left(\partial_\mu h_{\nu\lambda} + \partial_\nu h_{\lambda\mu} - \partial_\lambda h_{\mu\nu}\right), \tag{2.118}$$

or after keeping the first-order h terms:

$$\Gamma^{\rho}_{\mu\nu} = \frac{1}{2}\eta^{\rho\lambda}\left(\partial_\mu h_{\nu\lambda} + \partial_\nu h_{\lambda\mu} - \partial_\lambda h_{\mu\nu}\right). \tag{2.119}$$

We then construct the Riemann curvature tensor from 2.51. To first-order in h, and therefore neglecting the terms in $\Gamma \times \Gamma$ which are of higher order, the Riemann curvature tensor reduces to:

$$R_{\mu\nu\rho\sigma} = \frac{1}{2}\left(\partial_\nu\partial_\rho h_{\mu\sigma} + \partial_\mu\partial_\sigma h_{\nu\rho} - \partial_\mu\partial_\rho h_{\nu\sigma} - \partial_\nu\partial_\sigma h_{\mu\rho}\right). \tag{2.120}$$

From 2.52, the Ricci tensor in linearized gravity is:

$$R_{\mu\nu} = R^{\lambda}_{\mu\lambda\nu} = \frac{1}{2}\left(\partial_\sigma\partial_\nu h^{\sigma}_{\mu} + \partial_\sigma\partial_\mu h^{\sigma}_{\nu} - \partial_\mu\partial_\nu h - \Box h_{\mu\nu}\right), \tag{2.121}$$

where h is the trace of the tensor $h^{\alpha\beta}$ with respect to the Minkowski metric:

$$h \equiv \eta^{\mu\nu}h_{\mu\nu}, \tag{2.122}$$

and where we have used the D'Alambertian operator denoted by $\Box$:

$$\Box \equiv -\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2}. \tag{2.123}$$

We contract the Ricci tensor to obtain the curvature scalar R:

$$R = \eta^{\mu\nu} R_{\mu\nu} = \partial_\mu \partial_\nu h^{\mu\nu} - \Box h. \tag{2.124}$$

We finally obtain the Einstein tensor, which gives at first-order:

$$\begin{aligned} G_{\mu\nu} &= R_{\mu\nu} - \frac{1}{2} R \eta_{\mu\nu} && (2.125) \\ &= \frac{1}{2} \left\{ \partial_\sigma \partial_\nu h^\sigma_\mu + \partial_\sigma \partial_\mu h^\sigma_\nu - \partial_\mu \partial_\nu h - \Box h_{\mu\nu} \right. \\ &\quad \left. - \eta_{\mu\nu} \partial_\sigma \partial_\lambda h^{\sigma\lambda} - \eta_{\mu\nu} \Box h \right\}. && (2.126) \end{aligned}$$

This expression can be simplified via the introduction of the trace-reversed metric $\overline{h}_{\mu\nu}$:

$$\overline{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2} h \eta_{\mu\nu}. \tag{2.127}$$

Since $\eta_{\mu\nu}\eta^{\mu\nu} = 4$, we have that the trace of $\overline{h}_{\mu\nu}$ is:

$$\eta^{\mu\nu} \overline{h}_{\mu\nu} = h - 2h = -h, \tag{2.128}$$

which explains why $\overline{h}_{\mu\nu}$ is called the trace-reversed metric perturbation. When substituting $h_{\mu\nu} = \overline{h}_{\mu\nu} + \frac{1}{2} h \eta_{\mu\nu}$ in equation (2.126), all second-order derivatives of h cancel out so that we are left with the following generalized expression for the linearized Einstein equation (with no cosmological constant):

$$\Box \overline{h}_{\mu\nu} + \eta_{\mu\nu} \partial^\rho \partial^\sigma \overline{h}_{\rho\sigma} - \partial^\rho \partial_\nu \overline{h}_{\mu\rho} - \partial^\rho \partial_\mu \overline{h}_{\nu\rho} = -\frac{16\pi G}{c^4} T_{\mu\nu}. \tag{2.129}$$

In fact, one can use a suitable change in coordinates to simplify this equation into a standard wave equation 2.141. To see this, we consider a general infinitesimal coordinate transformation, which can be written as:

$$x'^\mu = x^\mu + \xi^\mu (x). \tag{2.130}$$

Using the transformation law of the metric 2.30, we find that the transformation of $h_{\mu\nu}$ to the lowest order is:

$$h_{\mu\nu} \rightarrow h'_{\mu\nu} = h_{\mu\nu} - \left(\partial_\mu \xi_\nu + \partial_\nu \xi_\mu\right) \tag{2.131}$$

We can use the gauge freedom ξ^μ so as to choose a coordinate system that satisfies the so-called Lorentz gauge condition:

$$\partial^\mu \overline{h}'_{\mu\nu} = 0, \tag{2.132}$$

To prove this, let us express equation 2.131 in terms of the trace-reversed metric perturbation $\overline{h}_{\mu\nu}$:

$$\overline{h}_{\mu\nu} \rightarrow \overline{h}'_{\mu\nu} = \overline{h}_{\mu\nu} - \left(\partial_\mu \xi_\nu + \partial_\nu \xi_\mu - \eta_{\mu\nu} \partial_\rho \xi^\rho\right), \tag{2.133}$$

which implies that

$$\partial^\nu \overline{h}_{\mu\nu} \rightarrow \left(\partial^\nu \overline{h}_{\mu\nu}\right)' = \partial^\nu \overline{h}_{\mu\nu} - \Box \xi_\mu. \tag{2.134}$$

Starting with expressing this quantity

$$\partial^\nu \overline{h}_{\mu\nu}(x) = f_\mu(x) \tag{2.135}$$

for some function f_μ, we can obtain

$$\left(\partial^\nu \overline{h}_{\mu\nu}\right)' = 0 \tag{2.136}$$

by choosing ξ_μ so that

$$\Box \xi_\mu = f_\mu(x), \tag{2.137}$$

an equation that always admits a solution by virtue of the invertibility of the d'Alembertian operator. Denoting by $G(x)$ a Green's function of the d'Alembertian operator so that:

$$\Box_x G(x - y) = \delta^4(x - y), \tag{2.138}$$

we obtain an explicit expression for ξ^μ as a function of f_μ:

$$\xi^\mu(x) = \int d^4x G(x-y) f_\mu(y) \tag{2.139}$$

The advantage of working with a gauge that satisfies the Lorentz gauge condition is that it greatly simplifies the expression of the Einstein tensor, which reads

$$G_{\mu\nu} = \frac{1}{2}\Box\overline{h}_{\mu\nu} \tag{2.140}$$

so that Einstein equation eventually becomes a simple wave equation:

$$\Box\overline{h}_{\mu\nu} = -\frac{16\pi G}{c^4}T_{\mu\nu}. \tag{2.141}$$

Note that Lorentz gauge condition is a set of 4 separate conditions which reduces the 10 independent components of the symmetric metric perturbation $h_{\mu\nu}$ to 6 independent components. In fact we argue in the next section that there remain 4 additional gauge degrees of freedom that can be used in a specific choice of coordinates called the transverse-traceless gauge so as to impose 4 additional conditions on $h_{\mu\nu}$ so as to finally leave only 2 independent components of the metric perturbation.

The Transerve-Traceless Gauge Equation 2.141 is the basic equation allowing one to study gravitational waves within the context of a linearized gravitational field. To study the propagation of GWs as well as their interactions with detectors, we are interested in spacetime geometry outside the source ($T_{\mu\nu} = 0$). In vacuum, the linearized Einstein equation reduces to:

$$\Box\overline{h}_{\mu\nu} = 0. \tag{2.142}$$

Let us now observe that Lorentz gauge condition 2.132 allows for additional degrees of freedom. Starting from

$$\left(\partial^\nu\overline{h}_{\mu\nu}\right)' = \partial^\nu\overline{h}_{\mu\nu} - \Box\xi_\mu \tag{2.143}$$

we can perform an additional coordinate transformation

$$x^\mu \rightarrow x^\mu + \xi^\mu \tag{2.144}$$

by taking

$$\Box \xi_\mu = 0 \tag{2.145}$$

at the same time as we impose

$$\partial^\nu \overline{h}_{\mu\nu} = 0 \tag{2.146}$$

since this choice would still preserve the condition

$$\left(\partial^\nu \overline{h}_{\mu\nu}\right)' = 0. \tag{2.147}$$

As a result, from the 6 remaining independent components of $\overline{h}_{\mu\nu}$, which satisfies the linearized Einstein equation in vacuum 2.142, we can subtract functions $\xi_{\mu\nu}$ as long as they satisfy

$$\Box \xi_{\mu\nu} = 0. \tag{2.148}$$

We can define such functions as:

$$\xi_{\mu\nu} \equiv \partial_\mu \xi_\nu + \partial_\nu \xi_\mu - \eta_{\mu\nu} \partial_\rho \xi^\rho \tag{2.149}$$

since

$$\Box \xi_\mu = 0 \Rightarrow \Box \xi_{\mu\nu} = 0. \tag{2.150}$$

The functions $\xi_{\mu\nu}$ depend upon the 4 arbitrary independent functions ξ_μ, so we can always choose the gauge transformation $x^\mu \rightarrow x^\mu + \xi^\mu$ so as to impose 4 conditions on $\overline{h}_{\mu\nu}$. In particular, we can specialize the gauge to make it purely spatial:

$$\overline{h}_{tt} = \overline{h}_{ti} = 0 \tag{2.151}$$

and traceless:

$$\overline{h} = \overline{h}^{\mu}_{\mu} = 0. \tag{2.152}$$

In this case, the gauge condition implies that the spatial metric is transverse:

$$\partial_{\mu}\overline{h}_{\mu\nu} = 0. \tag{2.153}$$

After nailing down all the remaining degrees of freedom in the choice of coordinates, the resulting gauge is said to be traceless-transverse (TT), and denoted by $\overline{h}^{TT}_{\mu\nu}$. Note that since $\overline{h}^{TT}_{\mu\nu}$ is traceless, we have that $\overline{h}^{TT}_{\mu\nu} = h^{TT}_{\mu\nu}$. Transverse traceless gauges make it explicit that gravitational waves have two polarization components. Consider for example a wave propagating in the z direction. We then have that

$$h^{TT}_{\mu\nu} = h^{TT}_{\mu\nu}\left(t - z\right) \tag{2.154}$$

is a valid solution to the wave equation $\Box h^{TT}_{\mu\nu} = 0$. The Lorentz gauge condition $\partial_z h^{TT}_{\mu\nu} = 0$ then implies that $h^{TT}_{z\nu}\left(t - z\right)$ is a constant, which must be zero since $h^{TT}_{\mu\nu} \to 0$ as the distance to the source of gravitational waves becomes infinite. The only non-zero components of $h^{TT}_{\mu\nu}$ are then $h^{TT}_{xx}, h^{TT}_{xy}, h^{TT}_{yz}$ and h^{TT}_{yy}. By symmetry and because of the traceless condition, we finally have:

$$h^{TT}_{xx} = -h^{TT}_{yy} \equiv h_{+}\left(t - z\right) \tag{2.155}$$

$$h^{TT}_{xy} = h^{TT}_{yz} \equiv h_{\times}\left(t - z\right) \tag{2.156}$$

where the quantities $h_{+}\left(t - z\right)$ and $h_{\times}\left(t - z\right)$ are two independent waveforms known as

the "plus" and "cross" waveforms. In summary, we have in the traceless-transverse gauge:

$$\overline{h}_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Interactions of Gravitational Waves with Matter and Detectors

Let us again consider a plane gravitational waves moving along the $z-$axis. In the transverse traceless coordinate system $(x^\alpha) = (ct, x, y, z)$, the spacetime metric g satisfies:

$$g_{\mu\nu} dx^\mu dx^\nu = -c^2 dt^2 + \left(\delta_{ij} + h_{ij}^{TT}\right) dx^i dx^j. \tag{2.157}$$

If we consider a particle A moving freely in spacetime, with no force acting on it except for the influence of the gravitational waves, its geodesic equation 2.49 $x^\alpha = X_A^\alpha(\tau)$ satisfies:

$$\frac{d^2 X_A^\alpha}{d\tau^2} + \Gamma^\alpha_{\mu\nu} \frac{dX_A^\mu}{d\tau} \frac{dX_A^\nu}{d\tau} = 0, \tag{2.158}$$

where the Christoffel symbols are given in equation (2.119). For the metric perturbation in the TT gauge we have:

$$\Gamma^0_{0\alpha} = 0 \tag{2.159}$$

$$\Gamma^0_{ij} = \frac{1}{2c} \partial_t h_{ij}^{TT} \tag{2.160}$$

$$\Gamma^i_{00} = 0 \tag{2.161}$$

$$\Gamma^i_{0j} = \frac{1}{2c} \partial_t h_{ij}^{TT} \tag{2.162}$$

$$\Gamma^i_{jk} = \frac{1}{2c} \left(\partial_j h_{ik}^{TT} + \partial_k h_{ji}^{TT} - \partial_i h_{jk}^{TT}\right) \tag{2.163}$$

The $\alpha = i$ spatial component of the geodesic equation thus becomes:

$$\frac{d^2 X_A^\alpha}{d\tau^2} + \frac{1}{c}\partial_t h_{ij}^{TT} \frac{dX_A^0}{d\tau}\frac{dX_A^j}{d\tau} + \Gamma_{jk}^i \frac{dX_A^j}{d\tau}\frac{dX_A^k}{d\tau} = 0. \tag{2.164}$$

The only first-order term is the first one, so we are left with:

$$\frac{d^2 X_A^\alpha}{d\tau^2} = 0. \tag{2.165}$$

From the initial condition $X_A^\alpha(0) = 0$ and given that $\frac{dX_A^i}{d\tau} = 0$, we obtain that the test particle stays at the point $(x, y, z) = (0, 0, 0)$ in the TT gauge. That the test particle does not move in the TT coordinate system is a remarkable property of these coordinates (which intuitively means that they move with the waves), which does not imply that measured physical distances are not impacted by the passage of the gravitational wave. In fact the distance L between two freely falling test particles would oscillate at the passage of the gravitational wave. To see this, let us now consider two free-moving test particles A and B. At the proper time t_1 an observer located on a reference frame attached to A sends a light signal in direction to B, which is reflected back to A where it arrives at the proper time t_2. We define the distance between A and B as:

$$L = \frac{c}{2}(t_2 - t_1) \tag{2.166}$$

We now measure the variation δL that is caused by the passage of the gravitational wave. Assuming that A and B are infinitely close, we have that:

$$L^2 = g_{\mu\nu}\left(x_B^\mu - x_A^\mu\right)\left(x_B^\nu - x_B^\nu\right), \tag{2.167}$$

or equivalently:

$$L^2 = g_{ij}\left(x_B^i - x_A^i\right)\left(x_B^j - x_A^j\right) \tag{2.168}$$

since:

$$x_A^0 = ct = x_B^0. \tag{2.169}$$

Assuming that the test particle A is located at the origin $(x_A^x, x_A^y, x_A^z) = (0, 0, 0)$, we obtain:

$$L^2 = g_{ij} x_B^i x_B^j = \left(\delta_{ij} + h_{ij}^{TT}\right) x_B^i x_B^j. \tag{2.170}$$

If we denote by $\overrightarrow{n}$ the unit space vector joining A and B for the Minkowski metric, we have $x_B^i = L_0 n^i$ with:

$$L_0 = \delta_{ij} x_B^i x_B^j. \tag{2.171}$$

We finally obtain:

$$L = \sqrt{L_0^2 \left(\delta_{ij} + h_{ij}^{TT}\right) n^i n^j} = L_0 \sqrt{\left(1 + h_{ij}^{TT} n^i n^j\right)}. \tag{2.172}$$

Keeping only the first order terms in h, we have:

$$L = L_0 \left(1 + \frac{1}{2} h_{ij}^{TT} n^i n^j\right). \tag{2.173}$$

As a result, the relative change in measured length $\frac{\delta L}{L_0} \equiv \frac{L - L_0}{L_0}$ due to the impact of the gravitational wave is:

$$\frac{\delta L}{L_0} = \frac{1}{2} h_{ij}^{TT} n^i n^j. \tag{2.174}$$

Generation of Gravitational Waves by a Source

From equation 2.141, we have obtained the wave equation that relates the trace-reversed field $\overline{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2} h \eta_{\mu\nu}$ to the stress-energy tensor:

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial x_2^2} - \frac{\partial^2}{\partial x_3^2}\right) \overline{h}_{\mu\nu} = \frac{16\pi G}{c^4} T_{\mu\nu}. \tag{2.175}$$

In principle one can obtain the first-order contributions to the spatial component of the metric perturbation by solving this equation assuming that the source generates

a sufficiently weak gravitational field for the linearized version of Einstein equation to provide a satisfactory local approximation at least at some distance from the source. This leads to the so-called *quadrupole formula* that allows one to estimate the magnitude of the wave strain for a given source. In what follows, we present simple physical arguments, borrowed from the discussion in [80], that can be used to obtain an estimate for this strain. To do this, we define moments of the mass distribution for the object generating the gravitational waves, which we assume for simplicity to have no internal motion. The zeroth moment M_0 of the mass distribution, also known as monopole moment, is simply the mass itself:

$$M_0 = \int \rho d^3x = M, \tag{2.176}$$

where ρ denotes the mass density of the object. The first moment M_1 of the mass distribution, also known as dipole vector, is given by:

$$M_1 = \int \rho x_i d^3x = ML_i, \tag{2.177}$$

where L_i is a vector with the dimension of a length that describes the displacement of the center of mass from the origin (and as such is a frame-dependent quantity). The second moment M_2 of the mass distribution, also known as quadrupole moment tensor, is defined as:

$$M_2 = \int \rho x_i x_j d^3x = ML_{ij}, \tag{2.178}$$

where L_{ij} is a tensor with the dimension of a length squared. Using simple physical and dimensional arguments, it can be shown that M_0 and M_1 cannot contribute to the generation of GWs. To see this, let us start with M_0, and let us try to combine its value M with the distance to the source, r, in such a way as to produce a dimensionless wave strain h, which should fall off as $1/r$. Since G is expressed in $m^3 \times kg^{-1} \times s^{-2}$, c is expressed in $m \times s^{-1}$, M is expressed in kg, and r in m, we can check that the following

quantity:

$$h \sim \frac{G}{c^2}\frac{M}{r}. \tag{2.179}$$

is indeed dimensionless. Since the conservation of mass-energy implies that M is a constant for an isolated source, we cannot have a radiative strain h from the monopole moment. Consider now the second multipole moment. To obtain a dimensionless strain we need to take one time-derivative, which requires:

$$h \sim \frac{G}{rc^3}\frac{dM_1}{dt}. \tag{2.180}$$

where the extra factor of c converts the dimension of the time derivative to space, so that the whole expression is dimensionless. The time-derivative of the second monopole moment is:

$$\frac{dM_1}{dt} = \frac{d}{dt}\int \rho x_i d^3x = \int \rho \frac{dx_i}{dt} d^3x = P_i \tag{2.181}$$

which is the total momentum of the source, so we finally obtain:

$$h \sim \frac{G}{c^3}\frac{P}{r}. \tag{2.182}$$

Since the conservation of momentum implies that P is a constant for an isolated source, we cannot have a radiative strain h from the dipole moment either.[5] Extending this analysis, it can be seen that the first moment to contribute to the generation of gravitational waves is M_2, and it contributes to a strain wave given by:

$$h \sim \frac{G}{rc^4}\frac{d^2M_2}{dt^2}, \tag{2.183}$$

and there is no conservation principle that allows us to reject this term. We note that the factor $\frac{G}{c^4}$ is an extremely small quantity, which explains that only sources with very large accelerating or decelerating time-varying quadrupole moments are expected to gen-

[5] Even in situations when the conservation of mass-energy and momentum does not hold, M_0 and M_1 do not contribute to the generation of gravitational waves which are a rank 2 tensors.

erate gravitational waves of a measurable size. This discussion is consistent with the fact that gravitational interactions are very weak, and in fact the weakest of all known interactions. As a result, even extremely large energies radiated from gravitational waves will generate an extremely small strain. For example, in the case of GW150914 [9] the energy radiated by the merger of the two black holes was equivalent to as much as 3 solar masses having been dissipated under the form of gravitational waves while the peak GW strain detected was merely 10^{-21} m! Astrophysical objects that are natural candidates for the generation of detectable gravitational waves are compact objects evolving at high velocity. If we define the gravitational luminosity factor L_u as the total energy radiated by the gravitational wave per unit of time, we have:

$$L_u \sim \frac{c^5}{G} s_y^2 \left(\frac{R_S}{R}\right)^2 \left(\frac{v}{c}\right)^6, \tag{2.184}$$

where R is the spatial extension of the source, R_S its Schwarzschild radius, v its velocity and s_y an asymmetry factor ($s_y = 0$ for a symmetric source). For a compact object ($\frac{R_S}{R} \sim 1$), with a strong deviation from spherical symmetry ($s_y \sim 1$), and moving at relativistic speed ($\frac{v}{c} \sim 1$), the quantity of energy radiated can be extremely large:

$$L_u \sim \frac{c^5}{G} = 3.6 \times 10^{52} W, \tag{2.185}$$

which is 10^{26} greater than the luminosity radiated by the Sun in the electromagnetic domain.

2.2 Detectors of Gravitational Waves

There are a variety of approaches that can be used to detect gravitational wave signals. Starting with Earth-based instruments, there have been historically two types of GW detectors, namely resonant bar detectors and laser interferometers. In what follows, we provide a brief discussion of resonant bar detectors before providing a more detailed

analysis of laser interferometers, which are the most sensitive instruments in operation today. It should also be noted that *pulsar timing array* experiments, which consist in analyzing a set of millisecond pulsars to search for correlated signatures in pulses emitted by the pulsars as a function of the pulsars' angular separations, can also be used to detect extremely low frequency (between 1 nHz and 1 mHz) gravitational waves [118]. Finally, let us note that at the largest spatial scales up to the scale of the observable universe, gravitational waves can be observed by measuring the polarization of the cosmic background radiation [35] and isolating the component related to gravitational waves hidden behind the largest polarization signals generated by cosmological density fluctuations [96].

2.2.1 Resonant Detectors

The first type of instrument used for the detection of gravitational waves was a resonant bar, or Weber bar, a large, solid bar of metal isolated from outside vibrations. In 1966, Joseph Weber of the University of Maryland constructed a gravitational-wave detector that consisted of a cylinder of aluminum 2 meters long and 1 meter in diameter. The idea was that when a gravitational wave was incident on the bar at a specific frequency, it would excite the bar's resonant frequency and the bar would start to ring like a bell. This "ringing" frequency, also called the resonant frequency, was 1660 Hz for Weber's bars [21].

Given the low intensity of gravitational wave strains, any ringing of the bar will be too small to hear or even to detect using normal ways of measuring vibration. Instead, crystals that produce an electric voltage when stretched or compressed (called a piezoelectric crystal) were mounted around the bar. As a result, Weber's bar was isolated from seismic and electromagnetic disturbances and housed in a vacuum. However, even the vibration of the aluminum atoms in the bar due to their temperature created significant noise and limited how small of a gravitational wave they could detect. Ultimately, they were limited to a strain $\frac{\delta L}{L}$ of about 10^{-16}. By 1969 Weber thought that he may have detected gravitational waves with his bar detectors [173]. This claim was ultimately not accepted

for many reasons including that other groups were not able to reproduce his rate of detections. Weber eventually lost the financial support of the National Science Foundation after a disputed claim of the detection of gravitational waves from a supernova observed in February 1987 [43]. Nonetheless, this launched the era of gravitational wave detectors, as resonant mass detectors began operating in countries around the globe. Most of the large baseline interferometers were constructed in the 90s. Today, there remains a network of three detectors (Explorer, Nautilus, and Auriga) operated under the general coordination of the International Gravitational Event Collaboration [18]. Since the early days, there have been many advancements in using resonant-mass bars to detect gravitational waves. Most bars today are made from new aluminum alloys, are cryogenically cooled to reduce the noise from the bar's thermal vibrations, have mechanical means to amplify the vibration, and piezoelectric crystals have been replaced with even more sensitive motion sensors known as SQUIDs (for Superconducting Quantum Interference Device). Different shapes (like spheres) have also been used to increase sensitivity to gravitational waves coming from different directions because bars are most sensitive to gravitational wave directly above or below the bar [11]. However, resonant-mass gravitational-wave detectors remain only sensitive to narrow ranges of gravitational wave frequencies.

2.2.2 Laser Interferometers

In what follows, we first provide an overview of existing and planned laser interferometer gravitational wave observatories, before turning to a presentation of the technical aspects of laser interferometry applied to gravitational wave detection.

Existing Detectors

The concept of a laser interferometer gravitational wave detector has a long and interesting history for which we only provide a brief review here (we refer to [153] for more details). The first discussion of the idea of how laser interferometers could be applied to the detection of gravitational waves can be traced back to a pair of papers, a paper writ-

ten in 1971 by a team at Hughes Research Lab led by Robert Forward, a former graduate student of J. Weber, and a paper written in 1972 by Rainer Weiss from MIT [178]. In his paper, Weiss states that he had been inspired by a 1956 paper by a Polish Physicist F.A.E. Pirani [140] and he further states that he had realized several years before (while teaching an undergraduate seminar) that the newly developed lasers could turn Pirani's thought experiment into a practical detection methodology. Weiss also notes that the idea "has been independently discovered by Dr. Philip Chapman of the National Aeronautics and Space Administration, Houston." American (Australian born) Philip K. Chapman, who had earned a doctorate in Instrumentation at MIT's Department of Aeronautics and Astronautics before joining NASA as a scientist-astronaut, had in fact been involved in separate discussions with the two teams, and has been instrumental in connecting them. Building upon this early work, Kip Thorne initiated in 1968 theoretical efforts on gravitational waves and their sources at Caltech, which have eventually led to launch of the Laser Interferometer Gravitational-Wave Observatory (LIGO) project. In the context of this project the initial LIGO observatories were conceived, built, and operated by Caltech and MIT with a funding by the National Science Foundation (NSF). A similar project, the Virgo project, was approved in 1993 by the French CNRS and in 1994 by the Italian INFN, the two institutes at the origin of the experiment, and the construction of the detector started in 1996 at the Cascina site near Pisa, Italy.

The LIGO and Virgo detectors are the main existing Earth-based interferometers. LIGO (Laser Interferometer Gravitational-wave Observatory) consists of two operating interferometers: one four kilometer interferometer in Livingston, Louisiana, and another one in the LIGO facility at Hanford, Washington. The sites are separated by roughly 3,000 kilometers. The LIGO and Virgo detectors have been operated for several years until they were stopped to allow for a sensitivity improvement. Advanced LIGO has been operational since the summer 2015 and Advanced Virgo is scheduled to become operational by the end of 2017. Further observation runs will be halted to allow commissioning efforts to further improve the sensitivity, with the aim to achieve design sensitivity in 2021

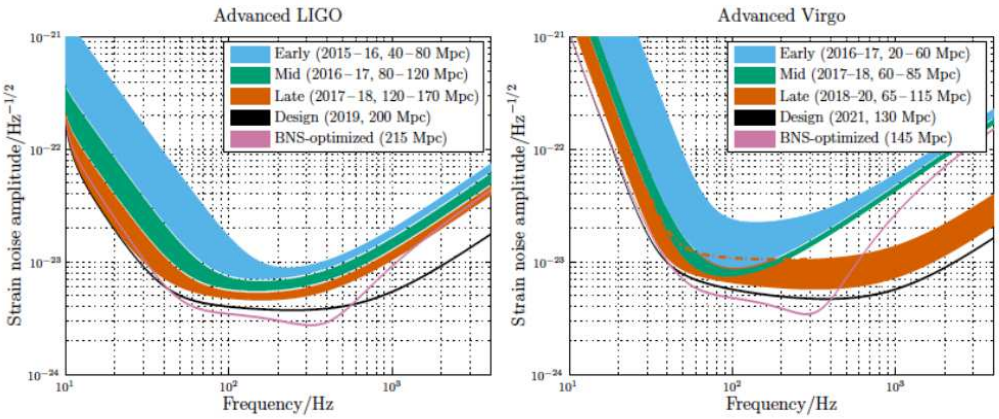

Figure 2.1: Projected sensitivity at later stages of development for Advanced LIGO and Advanced Virgo detectors.

(see Figure 2.1 for the scheduled sequence of improvements in the sensitivity of Advanced LIGO and Advanced VIRGO detectors). As a result of these improvements, in the most sensitive frequency region around 100 Hz, the design strain sensitivity is for Advanced LIGO a factor of 10 better than Initial LIGO. In addition, the low frequency end of the sensitivity band is moved from 40 Hz down to 10 Hz [3].

Other laser interferometers include GEO600, a six hundred meter interferometer constructed by a German-English collaboration near Hanover, Germany, as well as the Kamioka Gravitational Wave Detector (KAGRA) in Japan. KAGRA has two sets of 3kms arm length laser interferometric gravitational wave detectors which are being built in the tunnels of Kamioka mine and is now likely to enter operation in 2018 [113]. In the wake of LIGO's detection of gravitational waves, the Indian government has approved in the Spring 2016 the construction of the third LIGO interferometer (LIGO India).

Further Expected Upgrades

In spite of their early success, fundamental limitations at low frequency of the sensitivity of the 2nd generation detectors exist because of the presence of thermal noise of the suspension last stage and of the test masses, as well as seismic noise and related *gravity gradient noise*, also known as *Newtonian noise*. This Newtonian noise cannot be mitigated

by adding additional layers of seismic isolation because it is due to the actual gravitational attraction between the test masses and the density perturbations around the detector. A third-generation detector at the existing LIGO sites is being planned under the name *LIGO Voyager* to improve the sensitivity by an additional factor of two, and halve the low-frequency cutoff [124]. Additionally, a design exists for a larger facility with longer arms, which is called *Cosmic Explorer*. This project, with operation to commence post 2035, is an upgrade based on the LIGO Voyager technology, but expanded to a triangular configuration with up to 40 kms arms [124].

In parallel to these US projects, a third generation GW observatory project has also been developed in Europe to accommodate new infrastructures involving an underground site for the detector, to limit the effect of the seismic noise, and cryogenic facilities to cool down the mirrors to directly reduce the thermal vibration of the test masses. This project, known as the *Einstein Telescope* (ET) project, has been supported as Design Study by the European Commission under the Framework Programme 7 (FP7, Grant Agreement 211743). The Einstein Telescope, just like Cosmic Explorer, will also consist of three nested detectors, each composed of two interferometers with arms 10 kilometers long. One interferometer will detect low-frequency gravitational wave signals (2 to 40 Hz), while the other will detect the high-frequency components. The configuration is designed to allow the observatory to evolve by accommodating successive upgrades until reaching a sensitivity expected to be 100 times more sensitive than current instruments.

Finally, in addition to Earth-based interferometers, a Laser Interferometer Space Antenna (LISA) has been proposed as a joint NASA-ESA project, and is now a project led by ESA with the support of NASA. The goal is to develop and operate a space-based gravitational wave detector sensitive at frequencies between 10^{-4} Hz and 10^{-1} Hz. LISA is expected to detect gravitational-wave induced strains in space-time by measuring changes of the separation between fiducial masses in three spacecraft 5 million kilometers apart. As a first step, LISA Pathfinder was launched in December 2015 to test in flight the concept of low-frequency gravitational wave detection. It put two test masses

in a near-perfect gravitational free-fall, and controled and measured their motion with unprecedented accuracy. To do this, it used inertial sensors, a laser metrology system, a drag-free control system and an ultra-precise micro-propulsion system. The aims of the LISA Pathfinder mission is to (1) demonstrate, in a space environment, drag-free and attitude control in a spacecraft with two free proof masses, (2) test the feasibility of laser interferometry with picometer resolution at low frequency, and (3) test the endurance of the different instruments and hardware in the space environment. As a result of this experiment, it has been found that the two cubes at the heart of the spacecraft are falling freely through space under the influence of gravity alone, unperturbed by other external forces, to a precision substantially better than originally required. More precisely, the results demonstrate that two free-falling reference test masses, such as those needed for a space-based gravitational wave observatory like LISA, can be put in free fall with a relative acceleration noise with a square root of the power spectral density of $(0.54 \pm 0.01) \times 10^{-15} g/\text{Hz}$ for frequencies between 0.7 and 20 mHz; this value is lower than the LISA Pathfinder requirement by more than a factor 5 and within a factor 1.25 of the requirement for the LISA mission [22].

Overall, the different approaches and instruments that can be used to achieve GW detection should not be viewed as competing one with another. They actually target different sections of the GW spectrum (see Figure 2.2 borrowed from [40] for some typical amplitudes and wavelengths of gravitational waves across the GW spectrum, and the sensitivities of several detection methods).

While the advanced versions of LIGO and Virgo are designed to operate in the frequency range 10-10^3 Hz, it is expected that space-based detectors will cover a range between 10^{-4} Hz and 10^{-1} Hz. Turning to even lower frequencies, it is expected that pulsar timing arrays will also detect GWs around 10^{-8} Hz. Measures of the polarization of the cosmic background radiation will in turn allow us to probe gravitational events at frequencies around 10^{-16} Hz.

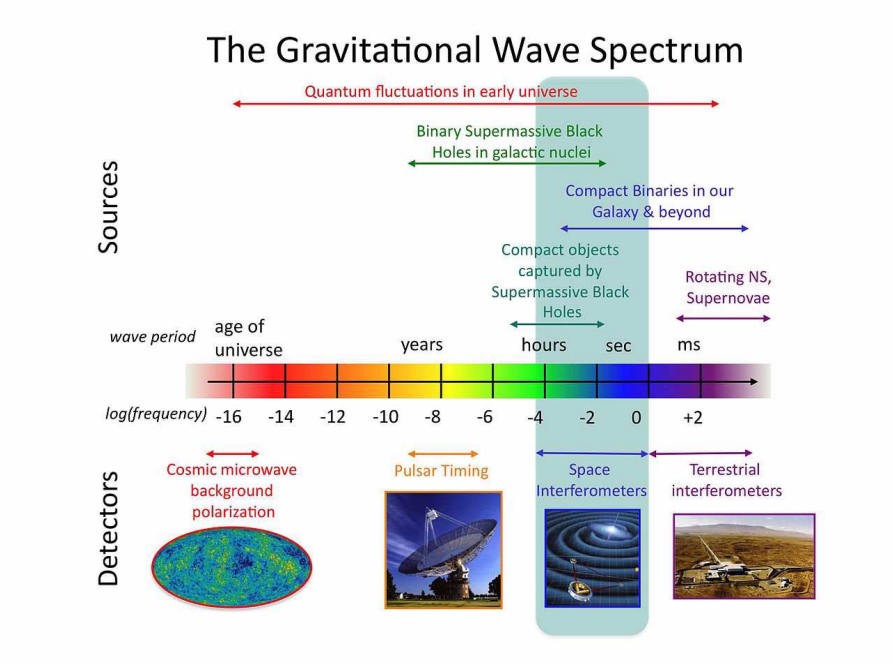

Figure 2.2: This diagram illustrates some typical amplitudes and wavelengths of gravitational waves across the entire spectrum, and the sensitivities of several detection methods. Figure borrowed from [40].

Laser Interferometer GW Observatory Technology

From a technical standpoint, a laser interferometer is typically made of a pair of L-shaped arms of a given length. To measure the relative lengths of the arms, a single laser beam is split at the intersection of the two arms. Half of the laser light is transmitted into one arm while the other half is reflected into the second arm. Mirrors are suspended as pendulas at the end of each arm and near the beam splitter (see Figure 2.3, borrowed from [156], for a diagram showing the components of the Advanced LIGO detectors). In order to artificially increase the arm length, one can bounce the light back and forth in the arms to increase the interaction time with the gravitational wave. With sufficiently large mirrors, one could construct a *Herriott delay line* [92] with hundreds of bounces [158]. It was subsequently proposed ([64], [63]) to instead use *Fabry-Perot optical resonators* in place of the delay lines. These cavities have the advantage of combining all of the many "bounces" of the delay line onto a single spot, thus greatly reducing the size, and thereby,

the cost, of the mirrors. Nearly all of the modern interferometers now use Fabry-Perot cavities instead of delay lines due to issues with scattered light in the latter [155]. The interferometer arm cavities are adjusted in length microscopically so that the fields from each arm interfere destructively at the Michelson anti-symmetric port. This causes almost all of the laser light to return towards the laser. By placing a partially transmitting mirror between the laser and the Michelson beamsplitter, this return light can be made to return towards the beamsplitter interfering constructively with the incoming laser light [10]. This so-called power recycling mirror is engineered to nearly equal the total scattering losses from the Michelson's optics [64] and thereby provide optimum power coupling from the laser source into the interferometer arms [7]. The modern GW interferometers with Fabry-Perot arm cavities have been able to increase the laser power on the beamsplitter by a factor of 65 by using this method [10]. The GEO600 detector has achieved a power gain of 1000 using power recycling [29].

If the lengths of both arms have remained unchanged, then the two combining light waves should completely subtract each other (destructive interference) and there will be no light observed at the output of the detector. However, if a gravitational wave were to slightly stretch one arm and compress the other, the two light beams would no longer completely subtract each other, yielding measurable light at the detector output. More specifically, interferometers measure the phase difference caused by passing GWs which stretch one arm as they compress the other. The gravitational wave strain amplitude is: $h = \frac{\delta L}{L}$. The advanced generation of LIGO-Virgo detectors is expected to be able to measure strains less than $h \simeq 10^{-21}$. If we consider that the freely falling particles are located at $z = 0$ and separated on the $x-$axis by the distance L_0, we obtain that the spatial distance between the two particles along the $x-$axis oscillates with the fractional change given by:

$$\frac{\delta L}{L_0} = \frac{1}{2} h_{xx}^{TT}\left(t, z = 0\right). \tag{2.186}$$

This quantity is particularly important because this is the quantity that it actually measured by interferometric gravitational wave detectors, where the change $\delta\phi$ in the photon

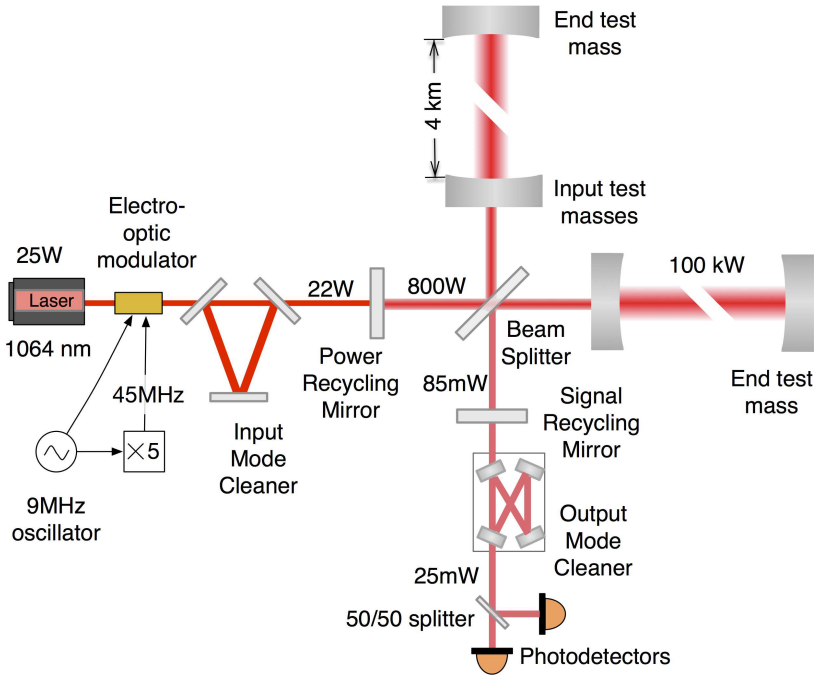

Figure 2.3: The Advanced LIGO detectors are Michelson interferometers with 4-km long arms. The arms contain Fabry-Perot optical cavities to amplify the signal from a gravitational wave. The electromagnetic field from the laser is modulated at radio frequencies (9 and 45 MHz) to generate signals used to control the detector, and the spatial profile of the beam is cleaned using an optical cavity (the Input Mode Cleaner). A mirror placed between the input mode cleaner and the beamsplitter (the Power Recycling Mirror) is used to recycle the laser power that is rereflected by the arms, and another mirror (the Signal Recycling Mirror) between the beamsplitter and the readout photodetectors is used to alter the frequency response of the interferometer. The Output Mode Cleaner removes excess laser light before the signal is measured by photodetectors. Picture borrowed from [156] (Figure 1), where more details can be found.

phase when travelling back and forth the arm of a laser interferometer in the presence of a gravitational wave is $\delta\phi = 4\pi\delta L/\lambda$. Here λ is the photon wavelength and L is the distance between the mirror located at the one end of each arm and the beam splitter at the other end of the arm. The magnitude h of the wave is often called the "wave strain".

Let us consider again for concreteness a gravitational wave propagating along the $z-$axis, and choosing Cartesian coordinates so that the interferometer's two arms lie along the x and y axis, with the beam splitter at the origin. Given that the only two non-zero components of the metric perturbation are:

$$h^{TT}_{xx} = -h^{TT}_{yy} \equiv h_+ \tag{2.187}$$

$$h^{TT}_{xy} = h^{TT}_{yz} \equiv h_\times \tag{2.188}$$

we have that the distance between the mirror at the end of each arm and the beam splitter varies over time according to:

$$\frac{\delta L_x}{L} = +\frac{1}{2}h_+ \tag{2.189}$$

$$\frac{\delta L_y}{L} = -\frac{1}{2}h_+ \tag{2.190}$$

These are precisely the changes in distance measured by laser interferometers. The gravitational wave sequentially squeezes one arm while the other one is stretched. In reality, it will not be the case that the source of the signal is ideally located so that the gravitational wave are generated in a direction that is exactly perpendicular to the plane formed by the two arms, meaning that a weighted combination of the two polarizations will impact each arm as a function of the exact location of the source in the sky.

2.3 Sources and Types of Gravitational Waves

The analysis in the previous section suggests that gravitational waves are very weak. The coupling coefficient $\frac{c^4}{8\pi G}$ in the Einstein equation 2.57 is a very large number, of order

10^{43}. This translates into the fact that if spacetime is elastic, with a geometry that is locally impacted by its mass-energy content, it is extremely stiff, which implies that the Newtonian law of gravitation is an excellent approximation to GR in weak gravitational fields. This also explains why gravitational waves have a small amplitude, even when their energy density is very high, and cannot be generated in the context of a laboratory experiment. Only very massive astrophysical objects travelling at relativistic speeds can be expected to generate gravitational waves sufficiently substantial to be noticeable. In fact the extreme amplitude of the strain at the point of coalescence of two 10 solar masses black holes is $h_{\max} \sim 10^{-16}$ located at the galactic centre [107]. The first observational validation of the existence of gravitational waves is the the observation by Taylor and Weisberg in 1982 [164] that the rate of decay of the orbit of the "Hulse-Taylor" PSR B1913+16 pulsar system, discovered by Hulse and Taylor in 1975 [98], matches GR prediction regarding the loss of energy of the system due to the emission of gravitational waves (see also [176] for more recent observations and related analyses). Binary pulsar systems are not the only sources of GWs, and one can classify the expected types of GW signals in 4 categories, with several astrophysical (or cosmological) phenomena that can generate detectable sources of GWs in each category:

- Inspiral GW signals,
- Periodic GW signals,
- Burst GW signals,
- Stochastic GW signals.

Another important distinction exists between continuous GW signals with a characteristic evolution time τ that is very long compared with the observation time T, and burst GW signals which are fast evolving gravitational waves signals with characteristic evolution time τ that is much smaller than the observation time T. If we choose T of the order of magnitude of 1 year as an observation time, sources that could be considered

continuous typically include compact binary systems in the gradual inspiral phase prior to the final few minutes of their evolution, and deformed neutron-stars. On the other hand, burst sources would include a core-collapse supernova, for which τ could be of order 10^{-3} seconds to a few minutes or late inspiral and coalescence stages of binary systems.

In what follows, we provide a brief description of the first three main types of GWs (inspiral, periodic and burst) and related astrophysical sources of GWs, ranked in increasing expected difficulty of detection [145]. We shall turn to a more detailed analysis of the stochastic GW background in the next chapter, where we make a key distinction between the SGWB of astrophysical origin and the SGWB of cosmological origin, and also discuss the distributional properties of this stochastic signal as well as standard detection methods.

2.3.1 Inspiral GW Signals

Inspiral GW signals, which are chirping signals produced by compact binary systems, are currently the best understood sources of GWs (see Figure 2.4 for a typical example of an inspiral gravitational wave signal). There are three main kinds of compact binary systems for this category of gravitational wave generators in the frequency range that can be accessible with Earth-bound detectors: Binary Neutron Star (neutron star-neutron star) or BNS systems, Binary Black Hole (black hole-black hole) or BBH systems, and Neutron Star-Black Hole Binary or NSBH systems.

To get a sense of the magnitude and frequency of the gravitational waves generated by compact binary systems, let us suppose that the binary system is well-separated so that each component can be treated as points. We let the masses be m_1 and m_2, and the orbital separation be R. We define $M = m_1 + m_2$ as the total mass and $\mu = \frac{m_1 m_2}{M}$ as the reduced mass of the system. The system of two masses m_1 and m_2 in a bound orbit is equivalently described by a single mass μ orbiting in an external potential determined by the total mass of the system M. In a general case, the mass moves in an elliptic orbit with eccentricity e and semi-major axis a. Here we assume circular orbits only so we

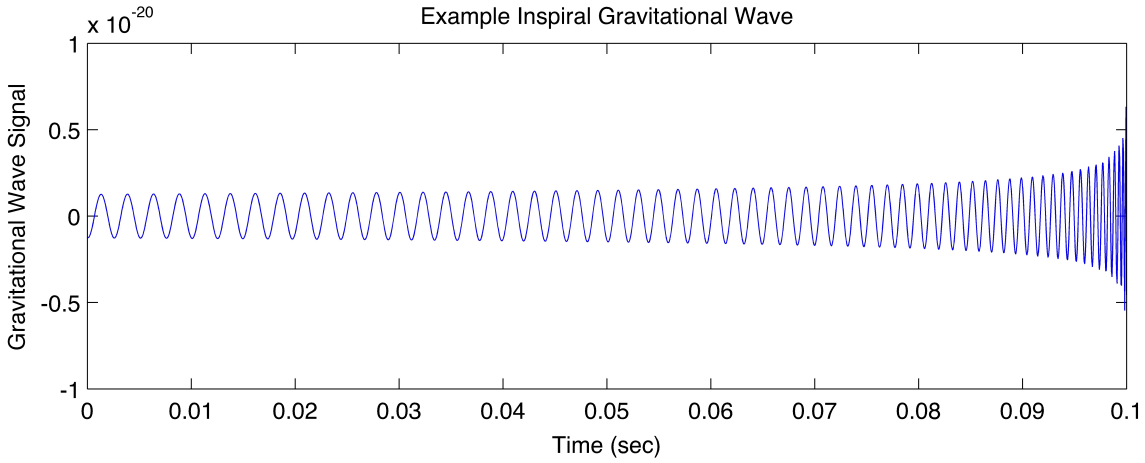

Figure 2.4: An example of signal from an inspiral gravitational wave source. Image taken from the "LIGO Science" website, available at http://www.ligo.org/science/GW-Inspiral.php.

have $e = 0$.[6] Using Kepler's third law, we can obtain the period P and angular orbital velocity Ω as a function of the total mass M and the semi-major axis a:

$$\frac{P^2}{a^3} = \frac{4\pi^2}{GM} \tag{2.191}$$

$$\Omega^2 = \frac{4\pi^2}{P^2} = \frac{GM}{a^3} \tag{2.192}$$

Approximating the mass quadrupole moment by $M_2 \sim \mu a^2$ and replacing $\frac{\partial^2}{\partial t^2}$ with Ω^2 in equation 2.183, we obtain that the amplitude at a distance $r >> R$ from this source is:

$$h \sim \left(\frac{G}{c^4}\right)\frac{1}{r}\left(\frac{\mu GM}{a}\right). \tag{2.193}$$

Using 2.192 to get rid of a in 2.193, and using that the frequency f of GW emission is

[6] Lowest order approximations for the GW strain when the binary orbit is eccentric are given in [138] and [139].

twice the orbital frequency $\frac{\Omega}{2\pi}$, that is $f = \frac{\Omega}{\pi}$, we finally obtain:

$$h = \sqrt{\frac{32}{5}} \frac{(GM_{ch})^{\frac{5}{3}}}{rc^4} (\pi f)^{\frac{2}{3}} \tag{2.194}$$

$$f = \frac{\Omega}{\pi} = \frac{1}{\pi} \left(\frac{GM}{a^3} \right)^{\frac{1}{2}}, \tag{2.195}$$

where we have defined the *chirp mass* M_{ch} via $M_{ch}^{5/3} = \mu M^{2/3}$ or

$$M_{ch} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}. \tag{2.196}$$

M_{ch} is named "chirp mass" because it is the quantity that determines how fast the binary "chirps" or "sweeps" through a frequency band. More generally, a chirp (or sweep) signal is a signal in which the frequency increases (up-chirp) or decreases (down-chirp) with time.

The construction of a generic model of the GW signal from coalescing black-hole binaries through their inspiral, merger and ringdown phases is an important requirement for GW detection. While the binary early inspiral can be modeled with analytic post-Newtonian calculations, the late inspiral and merger require numerical solutions of the full nonlinear Einstein equations (see [141] for pioneering work in this domain). These numerical relativity calculations involve at least seven parameters: the mass ratio of the binary (the total mass of the system is a scaling factor) and the components of each black hole's spin vector (see for example [91] for a simplified model of binary black holes coalescence). In addition to effective one-body calculations and the tools from numerical relativity, perturbation theory has proven to be an excellent approach for modeling the ringdown phase of general systems, especially, but not only, in situations with very large mass ratios [130], [56]. These methods allow us to obtain accurate templates for searching GWs generated by compact binary coalescence. In a non-spinning case with circular orbits, the inspiral source is described by 9 parameters, and 6 additional parameters are required for a spinning compact binary coalescence [167].

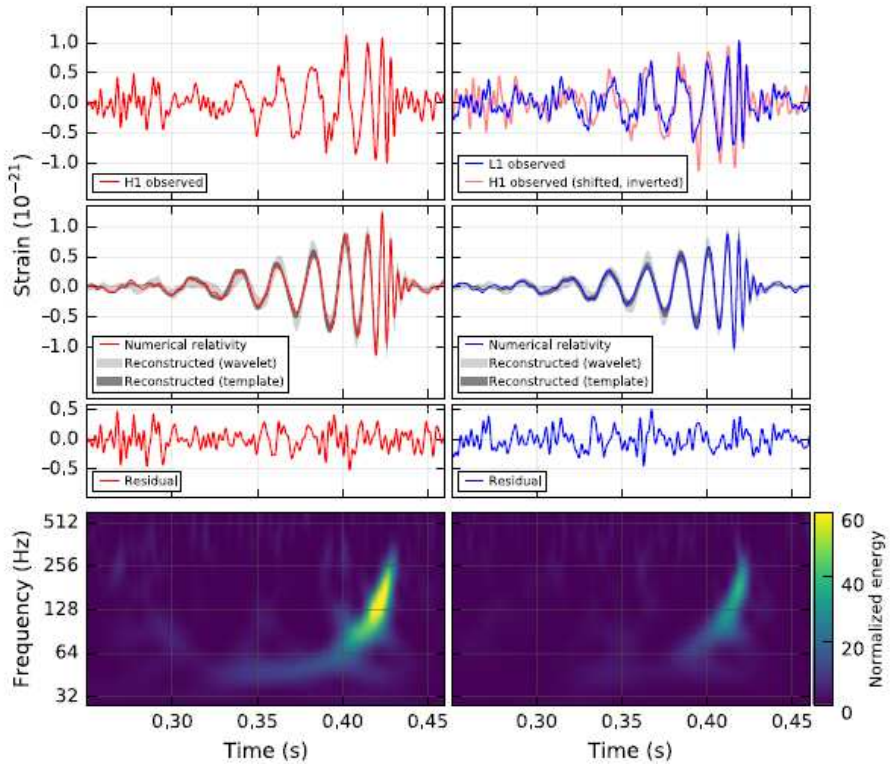

Figure 2.5: The gravitational-wave event GW150914 observed by the LIGO Hanford (H1, left column panels) and Livingston (L1, right column panels) detectors (Figure 1 in [9]).

Extrapolation from these observed binaries suggests that between a few to a few hundred binary neutron star mergers could be detected every year by existing GW detectors once they reach their target sensitivity. Models for the evolution of stellar populations also indicate that the measured rate of binaries containing black holes should also be relatively large, even though the uncertainties of population synthesis calculations are substantial. Our understanding of the population of binary black holes systems has been greatly enhanced by the first two detections by the twin detectors of the Laser Interferometer Gravitational-Wave Observatory (LIGO) that took place on September 14, 2015 and December 26, 2015. The first coincident signal, referred to as GW150914, is shown in Figure 2.5 below borrowed from [9].

The basic features of GW150914 point to it being produced by the coalescence of two black holes, i.e., their orbital inspiral and merger, and subsequent final black hole ringdown, as can be seen in Figure 2.6 below borrowed from [9]. The inferred initial

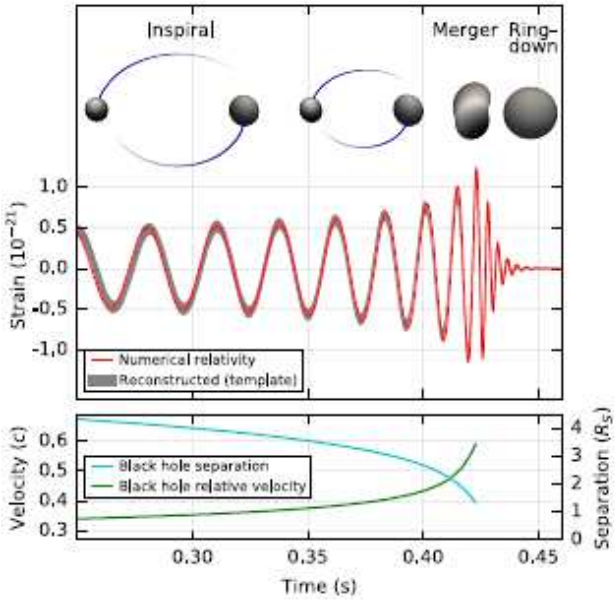

Figure 2.6: Effective black hole separation for GW150914 in units of Schwarzschild radii and the effective relative velocity given by the post-Newtonian parameter.

black hole masses are about 36.2 and 29.1 solar masses, and the resulting black hole mass is around 62.3, suggesting that about $3 = 36.2 + 29.1 - 62.3$ solar masses have been dissipated under the form of gravitational wave energy. Inspiral gravitational waves are generated during the end-of-life stage of binary systems where the two objects merge into one. As the two masses rotate around each other, their orbital distances decrease and their speeds increase, which causes the frequency of the gravitational waves to increase until the moment of coalescence.

The second detection of gravitational waves generated by the coalescence of a binary system of stellar mass black holes subsequently took place on December 26, 2015 ([5]). The GW151226 signal persisted in the LIGO frequency band for approximately 1 second, increasing in frequency and amplitude over about 55 cycles from 35 to 450 Hz, and reached a peak gravitational strain of about 3.4×10^{-22} meters. The inferred initial black hole masses are about 14.2 and 7.5 solar masses, and the resulting black hole mass is around 20.8. Another candidate, LVT151012, was found in the data with a low probability of being a false alarm (false alarm probability of about 2%), but not enough to claim

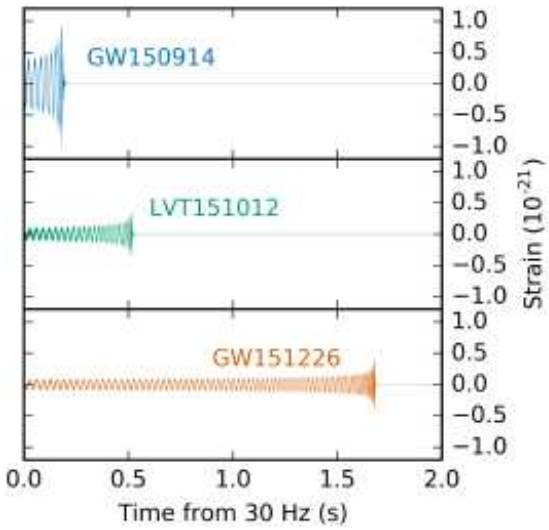

Figure 2.7: This Figure, borrowed from [45] (right panel of Figure 1), shows the time evolution of the 3 recovered signals from when they enter the detectors' sensitive band at 30 Hz.

a detection. If LVT151012 was to be regarded as an actual detection, the two black holes estimated masses would be 23 and 13 solar masses, respectively.[7] In Figure 2.7, borrowed from [45], we show the time evolution of the signals GW151226, GW150914 and LVT151012 from when they enter the detectors' sensitive band at 30 Hz. Finally, a third detection took place on January 4, 2017, in the second obervation run (O2). As was the case with the first two detections, the detected waves were generated by the coalescence of a binary system of stellar mass black holes. The black hole formed by the merger has a mass about 49 times that of our sun, which lies between the masses of the merged black holes form the first detections, with solar masses of approximately 62 (first detection) and 21 (second detection). Further detections expected from the new generation LIGO/Virgo detectors will help put tighter constraints on stellar population model parameters.

[7] LIGO detections are named "GW" followed by the date in YYMMDD format. LIGO/Virgo candidates start with a "LVT" for "LIGO-Virgo Trigger" followed by the date.

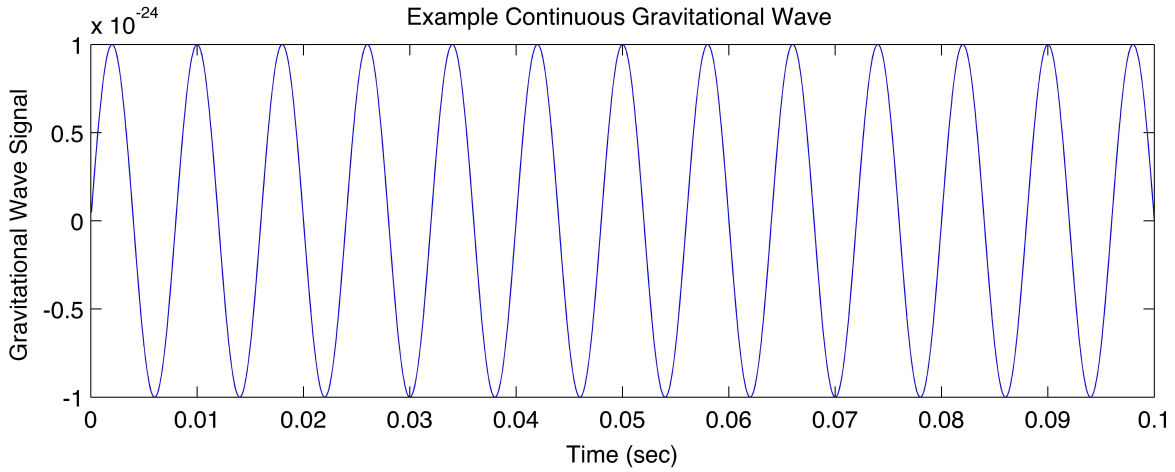

Figure 2.8: An example of a continuous gravitational wave source from a rotating neutron star. Image taken from the "LIGO Science" website, available at: http://www.ligo.org/science/GW-Continuous.php.

2.3.2 Periodic GW Signals

Periodic signals are continuous signals that are expected to have a duration much greater than the observation time and that are emitted at a nearly constant frequency. The prototypical source of a periodic GW signal (see Figure 2.8 for a stylized example) in the high-frequency band is a rotating non-axisymmetric neutron star. Neutron stars are compact remnants of supernovae having masses similar to the mass of the Sun, radii around 10 kms, which are supported by neutron degeneracy pressure. From an observational standpoint, the first pulsar was observed on November 28, 1967, by Jocelyn Bell Burnell and Antony Hewish [25] and found to emit in radio wavelengths. Pulsars have subsequently been found to emit in visible light, X-ray, and/or gamma ray wavelengths [86], and they are also expected to emit, under certain conditions discussed below, gravitational wave signals.

An asymmetry in a neutron star crust, caused, for example, by an oblateness that is misaligned with the star's spin axis (also known as a *mountain*), will radiate GWs with characteristic amplitude [170]:

$$h = \frac{4\pi^2 G I f^2}{c^4 r}\varepsilon, \tag{2.197}$$

were I is the star's moment of inertia, f is the wave frequency, r is the distance to the source, and ε is the dimensionless fractional distortion which characterizes the degree to which the star is distorted:

$$\varepsilon = (I_{xx} - I_{yy})/I, \tag{2.198}$$

where I_{xx} is the moment of inertia tensor.[8]

This "mountain" distortion could have been frozen into the crust or core of the star after it was born in the supernova, formed from material falling onto the star, or be produced and maintained though extremely large internal magnetic fields (larger even than the external fields described above). However, due to the huge gravitational field at the star surface the material forming the "mountain" needs to be really strong to not be flattened out. Rotational and tidal distortions can be described as a function of the internal structure of the stars. Various mechanisms have been proposed to explain how a neutron star can be distorted to give a value of ε that is interesting as a GW source, including distortion by accreting material from a companion star [170]. The accreting material comes from either the strong stellar wind in binaries with supergiant companions or the cirscumstellar disk in Be/X-ray binaries, and estimates of neutron star structure are consistent with a largest deformation $\varepsilon \sim 10^{-6}$ [146].

Searches for gravitational waves from a large selection of pulsars using data from the science runs of the initial generation of LIGO and Virgo did not lead to the detection of gravitational waves from any of 195 known pulsars for which data was collected. On the other hand, these results have produced improved estimates for the upper limits. For the Crab pulsar and Vela pulsar, less than about 1% and 10%, respectively, of their spin-down energy loss is due to gravitational radiation.[1]

[8] Even if the neutron star is axisymmetric, it can still emit GWs if the axis of symmetry of the star is not the same as the rotational axis [183].

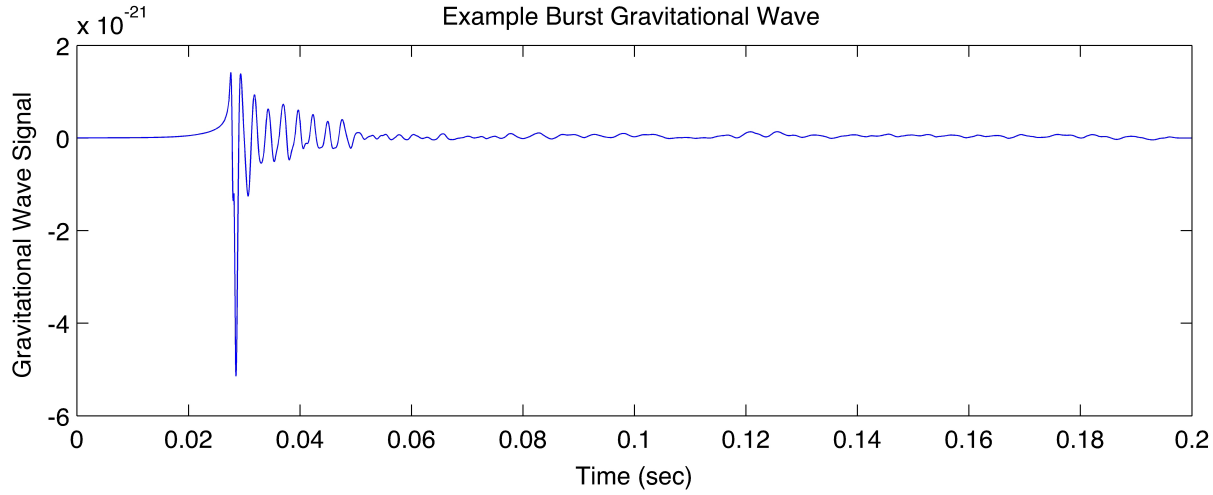

Figure 2.9: An example of signal from a burst gravitational wave source from a supernova core collapse. Image taken from the "LIGO Science" website, based on data from [133], available at http://www.ligo.org/science/GW-Burst.php.

2.3.3 Burst GW Signals

Gravitational waves produced in episodes that are comparatively short relative to the observation time are referred to as gravitational wave *bursts* (see Figure 2.9 for an example of a supernova core collapse gravitational wave signal would look like). According to this definition the coalescence of compact binaries including the late inspiral phase and the merger phase would qualify as a burst event. In practice, it is common to qualify as burst events that are much shorter in duration, typically less than 1 second. Thus, we regard gravitational wave (GW) bursts as *transient* signals with durations much shorter than the observational time scale and identifiable by a distinct arrival time.

A typical example of an astrophysical phenomenon that can generate a GW burst is the core collapse of a massive star such as type II supernovae. While the dynamic of the collapse of massive stars is relatively well understood from the theoretical standpoint [132], a number of uncertainties remain about the different core rotation parameters as well as the progenitor lower and upper mass cutoffs and may impact the process. As a result, the emitted wave form and available models for simulating core collapses

are subject to a high degree of uncertainty [51] . This stands in sharp contrast to the gravitational wave profile emanating from the coalescence of compact objects, which is, as recalled above, extremely well modelled.

Broadly speaking, type II supernovae result from the rapid collapse and violent explosion of a massive star, with a mass that is anywhere between around 8 to around 50 solar masses. The upper and lower bounds are defined as follows. Relatively light stars, with masses approximately lower than 8 solar masses, will turn into white dwarfs[9], which are compact stars supported by *electron degeneracy* pressure.[10] On the other hand, very massive stars with masses greater than about 50 solar masses are expected to collapse directly to black holes (*prompt collapse*) without producing supernovae.

When the compacted mass of the inert core exceeds the Chandrasekhar limit of about 1.4 solar masses, *electron degeneracy* is no longer sufficient to counter the gravitational compression and an implosion of the core takes place within seconds. Without the support of the imploded inner core, the outer core collapses inwards under gravity and reaches a velocity of up to about 20% of the speed of light and the sudden compression increases the temperature of the inner core to up to 100 billion kelvin. Core collapse in massive stars resulting in type II supernovae are expected to be an important source of GWs since they involve large amounts of mass flowing in a compact region (hundreds to thousands of kilometers) at relativistic speeds. On the other hand, a sufficiently high degree of asymmetry in collapse is required to generate strong gravitational waves, which may be the case if the star is rapidly rotating with a differential rotation caused by an asymmetric mass distribution during collapse [61]. Even without strong instabilities generated by rapid rotation, the asymmetric dynamics of core collapse is still likely to generate gravitational

[9] If the white dwarf is part of a binary system where there is mass transfer from the companion, *accretion-induced collapse* would still occur when the white dwarf mass eventually exceeds the Chandrasekhar limit of about 1.4 solar masses, and a Type Ia supernova is then produced.

[10] Electron degeneracy pressure is a particular manifestation of the more general phenomenon of quantum degeneracy pressure. The Pauli exclusion principle disallows two identical half-integer spin particles (electrons and all other fermions) from simultaneously occupying the same quantum state. The result is an emergent pressure against compression of matter into smaller volumes of space. Electron degeneracy pressure results from the same underlying mechanism that defines the electron orbital structure of matter and has been shown to explain the resistance of solid matter [65].

waves that could be detectable if emitted within the Local Group of galaxies.

Core collapses are not the only astrophysical phenomena that can produce burst GW signals. The ringdown phase of black hole coalescence for example can be regarded as generating short duration burst gravitational waves. The initial GW150914 detection was actually made by low-latency searches for generic gravitational wave transients [6] before it was recovered by matched-filter analyses that use relativistic models of compact binary waveforms. Of course, it may also be classified as the latest stage of a binary black hole coalescence and as such be also sorted in the inspiral signal category. More generally, one may expect to find burst gravitational waves from systems we never knew about before. As a result, the search for burst gravitational waves is more difficult compared to the search of inspiral or periodic signals since the lack of a proper understanding of the exact origin of the GWs implies that the detection analysis cannot be restricted to the use of a well-defined set of template waveforms.

3

Definition and Detection of the Stochastic Gravitational Wave Background

In contrast to aforementioned sources that produce GW signals that are bounded either in frequency (continuous gravitational waves) or in time (burst gravitational waves), the superposition of a sufficient number of sources that overlap both in the time and frequency domain results in an aggregate signal for which it is impossible to identify the marginal contribution of each source (*source confusion*). Countless GW sources from astrophysical origin, and also from fundamental cosmological processes, contribute to the generation of what is known as a gravitational wave *background*. Even if underlying signals are deterministic, the presence of uncertainty regarding the number of sources and uncertainty in the gravitational amplitude of each source implies that this aggregate signal is best modelled as a stochastic signal. We first discuss the possible astrophysical or cosmological origins of this stochastic gravitational wave background (SGWB), and then present the standard methods used in data processing and detection. We finally analyze its distributional properties with a particular emphasis on possible deviations from the Gaussian assumption.

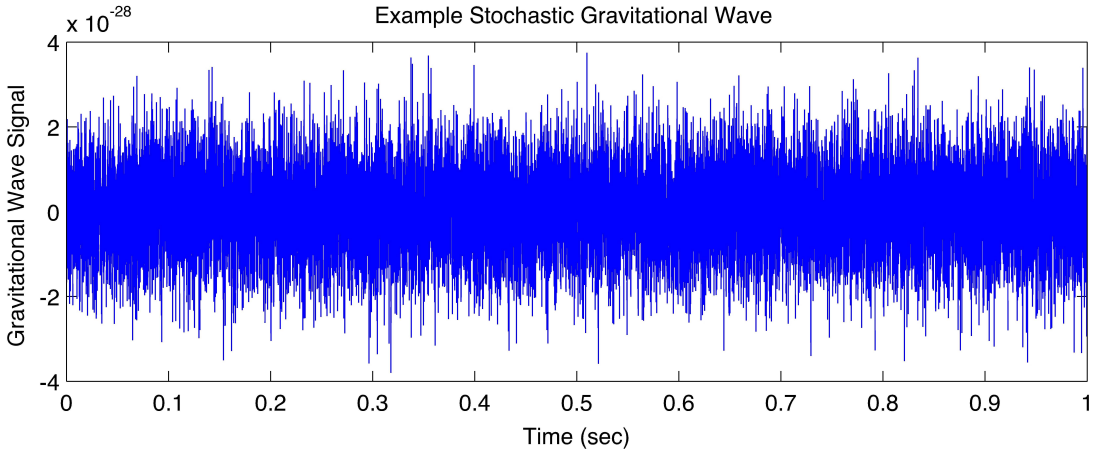

Figure 3.1: An example of stochastic gravitational wave background signal. Image taken from the "LIGO Science" website, available at http://www.ligo.org/science/GW-Stochastic.php.

3.1 Definition and Origins of Stochastic Gravitational Wave Backgrounds

At the intuitive level, the total gravitational wave signal received by a detector can be decomposed into the following components: the resolved signals and the stochastic gravitational wave background. The stochastic gravitational wave background can itself be decomposed into a cosmological gravitational wave background, and an astrophysical gravitational wave background generated from the superposition of a large number of unresolved signals, which results in a seemingly random signal (see Figure 3.1 for a stylized example of what a stochastic gravitational wave signal might look like).

3.1.1 Stochastic Gravitational Wave Background of Astrophysical Origin

As previously indicated, the astrophysical SGWB is formed by the superposition of a sufficient number of aforementioned overlapping sources of astrophysical origin such as inspiral signals from compact binary coalescences, periodic signals from rotating neutron

stars or burst signals from supernovae core collapses. The concept of overlapping sources can be formally defined in terms of a quantity known as *duty cycle* (DC). For a given type of astrophysical source of gravitational waves, the duty cycle measures the average fraction of the observation period T period for which the signal from the given type of sources is expected to be present. For a given source a value $DC = 1$ implies that a continuous signal emanating from one source is present over the whole measurement period, while a value $DC = 0.5$ implies that the signal is on average present only 50% of the time. Mathematically, the duty cycle is given by the following expression:

$$DC = \frac{1}{T} \int_0^{z_c} (1+z)\, \tau R(z)\, dz, \tag{3.1}$$

where $R(z)$ is the event rate observed in the Earth reference frame as a function of the redshift z (see 2.91), where the integration limit z_c corresponds to the epoch when the events of the given type first started, and where the characteristic evolution time τ is time dilated to $(1+z)\tau$ by cosmological expansion. Using this definition, the astrophysical gravitational wave background can itself be decomposed into a contribution from continuous sources with a duty cycle much greater than 1 ($DC >> 1$) and a contribution from transient sources with a duty cycle lower than 1 ($DC < 1$). The contribution from transient sources can itself be decomposed into a "shot" transient contribution when the duty cycle is much lower than 1 ($DC << 1$), and a "popcorn" transient contribution when the duty cycle is close to 1 ($DC \sim 1$) (see Figure 3.2 borrowed from [145] for an illustration of shot and popcorn SGWB signals).

In a nutshell, the discussion above can be summarized in terms of the following de-

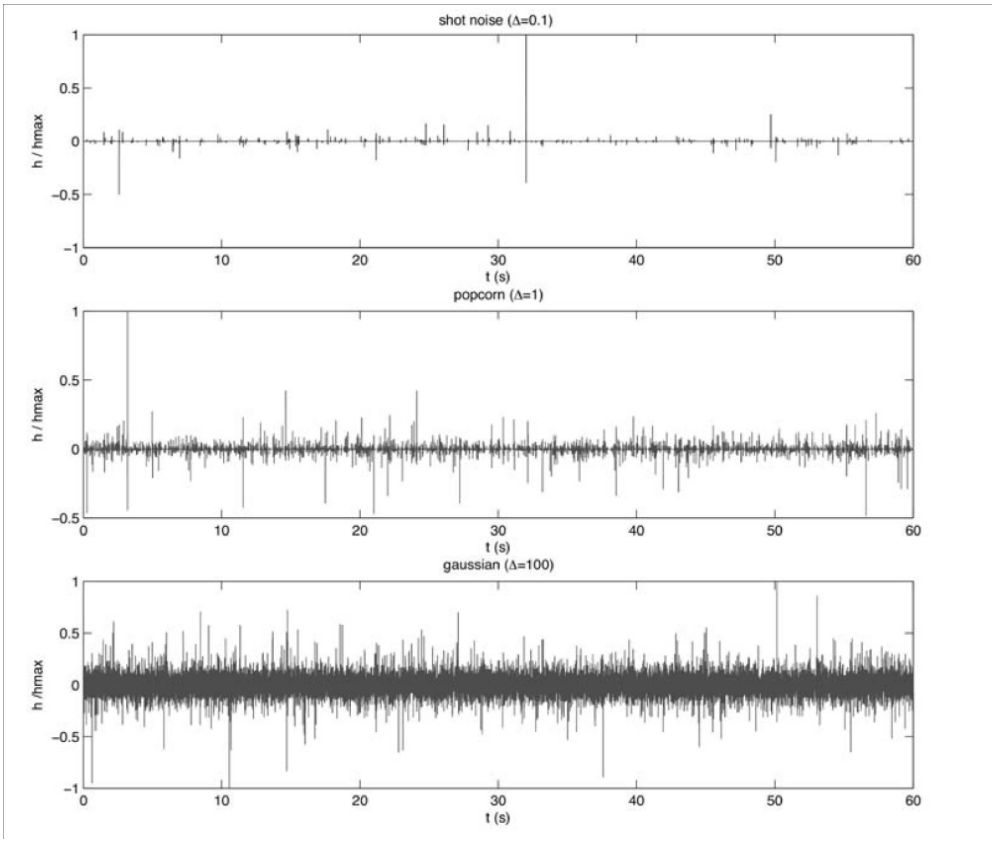

Figure 3.2: Time series corresponding to shot, popcorn and gaussian regimes - Figure borrowed from ([145]).

compositions:

$$\underbrace{S}_{\text{Aggregate signal}} = \underbrace{RS}_{\text{Resolved signal}} + \underbrace{SGWV}_{\text{Stochastic background}}$$

$$\underbrace{SGWV}_{\text{Stochastic background}} = \underbrace{CGB}_{\text{Background of cosmological origin}} + \underbrace{AGB}_{\text{Background of astrophysical origin}}$$

$$\underbrace{AGB}_{\text{Background of astrophysical origin}} = \underbrace{CAGB}_{\text{Contribution from continuous sources}} + \underbrace{BAGB}_{\text{Contribution from transient sources}}$$

$$\underbrace{BAGB}_{\text{Contribution from transient sources}} = \underbrace{SBAGB}_{\text{Contribution from "shot" transient sources}} + \underbrace{PBAGB}_{\text{Contribution from "popcorn" transient sources}}$$

Note that the resolved signals correspond to non-overlapping burst sources that fall above the detection threshold. In other words, if we could have access to perfect detectors with an idealized infinite sensitivity, then there would not be any non-overlapping burst contribution to the stochastic gravitational wave background, but the stochastic gravitational wave background would still have (in addition to a contribution from cosmological origin) a component from an astrophysical origin, which would be generated by the superposition of continuous or burst signals that would be overlapping in the time and/or frequency domains.

Interestingly, the recent detections by LIGO have led the community to revise their expectations regarding the SGWB from binary black-holes. In particular, the detection GW150914 of inspiral and merger of two black holes with masses around 30 solar masses suggests a population of binary black holes with masses higher than originally expected, and therefore a contribution to the SGWB also higher than previously expected. In the most sensitive part of the Advanced LIGO/Virgo band for stochastic backgrounds around 25 Hz, and using the properties of GW150914, the energy density $\Omega_{\rm GW}(f = 25Hz)$ is predicted to take on a value equal to $1.1^{+2.7}_{-0.9}\times10^{-9}$ with 90% confidence [8]. This predicted value potentially allows for a measurement by the Advanced LIGO/Virgo detectors operating at their projected final sensitivity levels.

3.1.2 Stochastic Gravitational Wave Background of Cosmological Origin

In addition to the GWs of astrophysical origin, we also expect a stochastic background of gravitational waves to have been generated during the early evolution of the universe. Several cosmological scenarios imply the production of such primordial GWs. In what follows, we propose a brief overview of these cosmological sources of GWs and we refer to [27] as well as the references therein for more details.

The first scenario, consistent with the standard cosmological model, involves the amplification of quantum fluctuations of the inflation field in slowly rolling models of inflation [87], [163], [88]. Amplifications of vacuum fluctuations at the transition between the de Sitter, radiation dominated (RD) and matter dominated (MD) eras are expected to produce a GW background with characteristics that depend strongly on the fluctuation power spectrum developed during the early inflationary period [89]. In the most realistic scenario of "slow roll down" inflation, the inflaton field rolls toward the minimum of its potential, producing an acceleration of the expansion, while GWs are produced by fluctuations that go out the Hubble radius during inflation, and re-enter at the radiation era. The resulting spectrum has a f^{n_T} dependency, where n_T is a negative number with a small absolute value, perhaps of the order of 10^{-3} [120], which translates into GWs that may be out of the detection range given realistic estimates for expected improvements of detector sensitivities.

The second scenario, also consistent with the standard cosmological model [33], [117], involves first order phase transition in the early stages of the universe evolution [37], [36], [38]. In the early stages of its evolution, the Universe may have undergone several episodes of phase transitions, in which the symmetry of particle-physics fundamental interactions spontaneously broke. This may occur for instance at the quantum chromodynamics (QCD) scale (150 Mev) and electroweak scale (100 GeV) or even earlier, at the grand unified scale (see [120] and references therein). The standard model predicts a smooth

crossover, but in its supersymmetric extensions, the transition from a metastable phase (the false vacuum) to the state of broken symmetry (the true vacuum) may have generated a large amount of GWs that could be produced when bubbles of the new phase are nucleated, grow and collide at very high velocities as they become more numerous.

The third scenario, which has received a lot of attention in the literature, involves the presence of primordial GWs emitted by *cosmic strings*, which are one-dimensional line-like topological defects formed during symmetry breaking phase transitions in the very early universe [109] as well as in later stages of the universe evolution. In addition to cosmic strings which existence is predicted by a number of field theories, so called "super strings" may have been produced in string theory inspired inflation scenarios [34], [106] in M-theory models. The main differences between cosmic strings and super strings are that (1) the latter reconnect when they meet with probabilities p that can be less than 1 (the values suggested for p are in the range $[10^{-3}; 1]$[100]), and that (2) more than one kind of string can form.[1] The effect of a small reconnection probability is to increase the time it takes for the network to reach equilibrium, and to increase the density of strings at equilibrium [159]. Although the analysis of cosmic microwave background data has led to rule out cosmic strings as the likely primary source of density perturbations[2], they are still believed to be at the origin of a number of interesting phenomena, including precisely the generation of primordial gravitational waves [172]. Cosmic strings originating in the symmetry breaking of a Grand Unified Theory possess an enormous mass per unit length μ ($G\mu \simeq 10^{-6}$, where G is Newton's constant). After formation the network of strings quickly evolves towards an attractor solution called the "scaling regime" (see [172] as well as references therein) where the energy density in long strings (also known as infinite strings) remains a constant fraction of the total background energy density. This is achieved by intercommutations and self-intersections of the strings leading to the production of small loops, which then decay by emitting gravitational radiation. In

[1] Cosmic superstrings may form also Y-junctions where 3 different strings meet.

[2] While most inflationary models produce Gaussian random phase initial conditions, defect models produce non-Gaussian perturbations particularly on small scales [23].

this way, some of the energy input into the string network coming from the stretching of the strings due to the expansion of the universe is transferred to the background. Despite this superimposed small-scale structure due to intercommutations, long strings are almost straight over distances of the order of the horizon, so that the scaling solution can be pictured as having a fixed number of long strings per Hubble volume at any given time [129]. From the observational standpoint, the presence of very large mass-per-unit-length strings is not supported by CMB observations [105], but it is still conceivable that strings of lower energy scale may contribute to the generation of GWs. Until recently, it appeared that the gravitational effects of strings were too weak to be observable. However, it has been shown in [57] and [58] that GW bursts emitted from cusps of oscillating loops could be detectable by ground-based interferometers as a popcorn-like contribution. These strong bursts of gravitational radiation can potentially be detected individually and subtracted from the stochastic background [26], [58]. It has also been shown that LISA could detect a background from strings as light as $G\mu/c^2 = 10^{-16}$ [59]. Note that most of the literature has Focused on gravitational waves generated from two processes: cusps (whereby a segment of the string momentarily moves at the speed of light) and kinks (formed after two cosmic strings collide and reconnect). Gravitational waves can also be produced by cosmic superstring reconnections, but recent estimates of the burst amplitude imply that neither bursts nor the stochastic background would be detectable by Advanced LIGO, even for the most optimistic values of the reconnection probability and loop sizes [101].

3.2 Standard Detection Methods for Stochastic Gravitational Wave Backgrounds

Statistical inference method for GW data analysis can be used to answer two types of related questions, the detection question and the estimation question. The detection question is: "Is a gravitational-wave signal present in the data?". The estimation ques-

tion (assuming a positive answer to the detection question) is: "What are the physical characteristics, typically represented by a finite set of parameters, of the source?".

Gravitational-wave data analysis methods are typically divided into three main categories: (i) well-modeled deterministic signals, such as those from compact binary inspirals, for which we have waveform templates that are analyzed using a *matched-filter statistic* (see for example section 3.4 in [103]); (ii) poorly-modeled deterministic signals, such as those from core-collapse supernovae, which are analyzed in terms of an *excess power statistic* (see section 7.4 in [53]); and (iii) stochastic signals, such as those generated by cosmic strings, which are analyzed using a *cross-correlation statistic* between pairs of detectors.

A discussion of the methods used in the first 2 categories is beyond the scope of this thesis, which has a main focus on the stochastic gravitational wave background, and we refer the reader to the aforementioned references for more details. Broadly speaking, the optimal detection strategy to search for a stochastic background is to cross correlate the output of two detectors (or of a network of detectors) to eliminate the instrumental noise. In what follows, we focus on (iii) and propose a detailed overview of the standard cross-correlation approach used for detection of stochastic gravitational wave backgrounds. These methods can be classified into two main categories which are respectively known as the frequentist approach and the Bayesian approach. In this section, we discuss the application of these two approaches to data analysis (detection and estimation) for SGWB signals, and we also provide some elements of comparison of the two within the context of a unified framework [48].

3.2.1 Frequentist Approach to SGWB Data Analysis

A stochastic background of gravitational waves will generate a signal that is expected to be a weak random addition to the detector noise, and as such extremely difficult to isolate from the detector noise if a single detector is used. Fortunately, the extraction of the SGWB is made possible by the use of two, or more, independent detectors. Intuitively,

the presence of a SGWB will effectively introduce a correlated source of noise in the detection, which can be distinguished from the sources of detector noise if these are assumed to be sufficiently uncorrelated across the detectors. This is achieved by cross-correlating the measurement outputs from the various detectors, as will be discussed below. We first present the methodology in the simple case of two colocated and coaligned detectors with a cross-correlation analysis presented in the time-domain. An equivalent frequency-domain analysis can be obtained, which is related to the time-analysis by a simple Fourier transform, and which is a perspective that we favor when we discuss the *overlap reduction function* that is needed to handle the general situation where the detectors are not colocated and/or coaligned.

Case of Colocated and Coaligned Detectors

Consider for simplicity two gravitational wave detectors. The output of each detector is a collection of dimensionless strain measurements. Suppose that N such measurements are made with each detector at regular time intervals. Denote these measurements by a $T \times 2$ matrix h with components h_t^i, where $i = 1, 2$ labels the detector, and $t = 1, 2, ..., N$ is the discrete time of the measurement. We first decompose the measurement output for detector i in terms of noise versus signal, which gives when written in terms of random variables:

$$\mathcal{H}_i = \mathcal{N}_i + \mathcal{S}, \tag{3.2}$$

where $\mathcal{N}_i$ denotes the noise detected by the detector i and $\mathcal{S}$ denotes the signal detected by both detectors i (since we assume that the detectors are coincident and coaligned, i.,e., that they have identical location and arm orientations, the detectors receive the exact same signal so we drop the subscript i), and $\mathcal{H}_i$ is the total measurement for the detector i. We also assume that the detectors are identical, which implies that the noise for both detectors is drawn from the same distribution. We will generally assume that the noise has zero mean for both detectors, but the variance can be different across detectors (see chapter 5 for a detailed analysis of the implications of heterogenous detector sensitivities),

so we maintain the subscript i when needed.

In terms of the realization of such random variables for either one of the two detectors:

$$h_{it} = n_{it} + s_t. \tag{3.3}$$

So far, we have assumed that the gravitational wave signal is purely stochastic, with no contributions from resolved sources, either because such sources do not exist, or because they have been detected and subtracted from the total measurement output. One could extend this analysis to account for the presence of individual resolved sources of gravitational waves by using the following more general decomposition:

$$h_{it} = n_{it} + s_t + \sum_{k \leq K} b_t^k, \tag{3.4}$$

where b_t^k denotes the k^{th} signal generated by a resolved source of astrophysical origin, and K denotes the stochastic total number of bursts detected at a given point in time t. Since the focus in this section is purely on unresolved stochastic background, we take $K = 0$ for the moment, with the convention $b_t^0 \equiv 0$.

We denote by f_n and f_s, respectively, the density function for the noise and the signal distributions. (Here the subscript n or s is not an index - it stands for *noise* and *signal*, respectively). In other words, we have that :

$$\Pr\left(\mathcal{N} \in [n, n + dn]\right) = f_n\left(n\right) dn \tag{3.5}$$

$$\Pr\left(\mathcal{S} \in [s, s + ds]\right) = f_s\left(s\right) ds \tag{3.6}$$

Assuming that both the noise and signal are normally distributed, assuming a zero mean for all distributions, and denoting by σ_1 and σ_2, respectively, the standard deviations of the noise distributions, and by α the standard deviation of the signal distribution,

we obtain:

$$f_{n_i}(n_{it}) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{n_{it}^2}{2\sigma_i^2}} \text{ for } i = 1, 2 \tag{3.7}$$

$$f_s(s_t) = \frac{1}{\sqrt{2\pi}\alpha} e^{-\frac{s_t^2}{2\alpha^2}}, \tag{3.8}$$

where we also assume that both the noise and signal are weakly stationary processes so that their moments are constant through time. We further assume the noise in detector one and two are uncorrelated. Under these assumptions, we have:

$$f_n(n_{1t}, n_{2t}) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{n_{1t}^2}{2\sigma_1^2} - \frac{n_{2t}^2}{2\sigma_2^2}}, \tag{3.9}$$

and finally, assuming zero serial correlation:

$$f_n \equiv f_n(n_{1t}, n_{2t})_{t=1,\ldots,T} = \prod_{t=1}^{T} \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{n_{1t}^2}{2\sigma_1^2} - \frac{n_{2t}^2}{2\sigma_2^2}} \tag{3.10}$$

$$f_s \equiv f_s(s_t)_{t=1,\ldots,T} = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}\alpha} e^{-\frac{s_t^2}{2\alpha^2}}. \tag{3.11}$$

In the Gaussian case, the only unknown parameters are therefore α, σ_1 and σ_2. The standard Bayesian approach for signal detection consists in finding the value for the unknown parameters so as to minimize the false dismissal probability at a fixed value of the false alarm probability (see equations 4.85 and 4.86 for formal definitions). This decision rule is defined in terms of the so-called likelihood ratio Λ given by:

$$\Lambda = \frac{p_h|_{\mathbf{H}_1}}{p_h|_{\mathbf{H}_0}}, \tag{3.12}$$

where $p_h|_{\mathbf{H}_1}$ (respectively, $p_h|_{\mathbf{H}_0}$) is the conditional density for the measurement output if a signal is present (respectively, absent). Here we have for a measurement data point i that $p_{h_i}|_{\mathbf{H}_1} = f_n(h_i - s_i) f_s(s_i)$ and $p_{h_i}|_{\mathbf{H}_0} = f_n(h_i)$. In other words, the noise model defines the likelihood function since we demand that the residuals ($h_i - s_i$ when a signal

is present, and h_i otherwise) are consistent with the probability distribution of the noise. This criterion, known as the Neyman-Pearson criterion, is the standard frequentist approach to hypothesis testing. The key difference between the frequentist approach and the Bayesian approach for hypothesis testing (see next sub-section) is that in the latter approach a known prior probability for the relevant parameters, models and/or hypotheses is required. In any situation when the Bayesian framework cannot be applied because of the difficulty to assign probabilities to priors, the Neyman-Pearson criterion provides a convenient and optimal decision rule [131] that only depends on sample data.

As demonstrated in [62] (see page 8, discussion before equation (2.19)), one can approximate the likelihood ratio by the maximum likelihood detection statistic defined by:

$$\Lambda_{ML} = \frac{\max\limits_{\alpha,\sigma_1,\sigma_2} \int f_s(s)\, f_n(h-s)\, ds}{\max\limits_{\sigma_1,\sigma_2} f_n(h)} \tag{3.13}$$

where the maximum for the numerator is taken over all values for the unknown signal and noise standard deviation parameters α, σ_1 and σ_2, while it is taken for the denominator over all possible values for the unknown noise standard deviation parameters σ_1 and σ_2.[3] Hence the denominator of equation (3.13) is given by:

$$\max_{\sigma_1,\sigma_2} f_n(h) = \max_{\sigma_1,\sigma_2} \prod_{t=1}^{T} \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{h_{1t}^2}{2\sigma_1^2}-\frac{h_{2t}^2}{2\sigma_2^2}} = \max_{\sigma_1,\sigma_2} \frac{1}{(2\pi\sigma_1\sigma_2)^T} \exp\left[-\sum_{t=1}^{T}\frac{h_{1t}^2}{2\sigma_1^2} - \sum_{t=1}^{T}\frac{h_{2t}^2}{2\sigma_2^2}\right]. \tag{3.14}$$

Introducing for $i = 1, 2$:

$$\overline{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T} h_{it}^2, \tag{3.15}$$

we finally have that:

$$\max_{\sigma_1,\sigma_2} f_n = \max_{\sigma_1,\sigma_2} \frac{1}{(2\pi\sigma_1\sigma_2)^T} \exp\left[-\frac{T}{2}\left(\frac{\overline{\sigma}_1^2}{\sigma_1^2} + \frac{\overline{\sigma}_2^2}{\sigma_2^2}\right)\right]. \tag{3.16}$$

[3] It should be noted that the values for σ_1 and σ_2 that maximize the numerator of Λ_{ML} (given in equation 3.40) are a priori different from the values that maximize the denominator (given in equation 3.15).

It is straightforward to see that the maximum for equation (3.16) is reached for $\sigma_i^2 = \overline{\sigma}_i^2$, and that this maximum is given by:

$$\max_{\sigma_1,\sigma_2} f_n = \frac{1}{\left(2\pi\overline{\sigma}_1\overline{\sigma}_2\right)^T} \exp\left[-\frac{T}{2}\left(1+1\right)\right] = \frac{1}{\left(2\pi\overline{\sigma}_1\overline{\sigma}_2\right)^T} \exp\left(-T\right). \tag{3.17}$$

Note that $\overline{\sigma}_i$ defined above is not equal to the standard unbiased sample estimator for the variance of the measurement made by detector i, which instead is given by:

$$\widehat{\sigma}_i^2 = \frac{1}{T-1}\sum_{t=1}^{T} h_{it}^2. \tag{3.18}$$

Of course the difference is negligible for large T values. Finally, we obtain the following expression:

$$\begin{aligned}
\Lambda_{ML}^G &= \frac{\max\limits_{\alpha,\sigma_1,\sigma_2} \int f_s\left(s\right) f_n\left(h-s\right) ds}{\max\limits_{\sigma_1,\sigma_2} f_n} && (3.19)\\
&= \left(2\pi\overline{\sigma}_1\overline{\sigma}_2\right)^T \exp\left(T\right) \\
&\quad \times \max_{\alpha,\sigma_1,\sigma_2} \prod_{t=1}^{T} \int_{-\infty}^{+\infty} f_s\left(s\right) \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{\left(h_{1t}-s_t\right)^2}{2\sigma_1^2} - \frac{\left(h_{2t}-s_t\right)^2}{2\sigma_2^2}\right] ds && (3.20)\\
&= \max_{\alpha,\sigma_1,\sigma_2} \prod_{t=1}^{T} \frac{\overline{\sigma}_1\overline{\sigma}_2}{\sigma_1\sigma_2} \int_{-\infty}^{+\infty} f_s\left(s_t\right) \exp\left[-\frac{\left(h_{1t}-s_t\right)^2}{2\sigma_1^2} - \frac{\left(h_{2t}-s_t\right)^2}{2\sigma_2^2} + 1\right] ds_t && (3.21)
\end{aligned}$$

where we use the subscript G in Λ_{ML}^G to emphasize that we are here in the Gaussian case. Maintaining this assumption of a Gaussian signal with a zero mean value, we obtain:

$$f_s \equiv f_s\left(s_t\right)_{t=1,\ldots,T} = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}\alpha} e^{-\frac{s_t^2}{2\alpha^2}} \tag{3.22}$$

and we thus have:

$$\begin{aligned}
\Lambda_{ML}^{G} &= \max_{\alpha,\sigma_1,\sigma_2} \prod_{t=1}^{T} \frac{\overline{\sigma}_1\overline{\sigma}_2}{\sqrt{2\pi}\alpha\sigma_1\sigma_2} \\
&\quad \times \int_{-\infty}^{+\infty} \exp\left[-\frac{s_t^2}{2\alpha^2} - \frac{(h_{1t}-s_t)^2}{2\sigma_1^2} - \frac{(h_{2t}-s_t)^2}{2\sigma_2^2} + 1\right] ds_t \qquad (3.23) \\
&= \max_{\alpha,\sigma_1,\sigma_2} \left\{ \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}\alpha} \frac{\overline{\sigma}_1\overline{\sigma}_2}{\sigma_1\sigma_2} \exp\left[-\frac{h_{1t}^2}{2\sigma_1^2} - \frac{h_{2t}^2}{2\sigma_2^2} + 1\right] \right. \\
&\quad \left. \times \int_{-\infty}^{+\infty} \exp\left[-\frac{s_t^2}{2\alpha^2} - \frac{s_t^2}{2\sigma_1^2} - \frac{s_t^2}{2\sigma_2^2} + \frac{s_t h_{1t}}{\sigma_1^2} + \frac{s_t h_{2t}}{\sigma_2^2}\right] ds_t \right\}. \qquad (3.24)
\end{aligned}$$

We define:

$$\begin{aligned}
A(s_t) &\equiv \exp\left[-\frac{s_t^2}{2\alpha^2} - \frac{s_t^2}{2\sigma_1^2} - \frac{s_t^2}{2\sigma_2^2} + \frac{s_t h_{1t}}{\sigma_1^2} + \frac{s_t h_{2t}}{\sigma_2^2}\right] \qquad (3.25) \\
&= \exp\left[-\frac{s_t^2}{2\sigma^2} + s_t\left(\frac{h_{1t}}{\sigma_1^2} + \frac{h_{2t}}{\sigma_2^2}\right)\right] \qquad (3.26)
\end{aligned}$$

with σ^2 such that:

$$\frac{1}{\sigma^2} = \frac{1}{\alpha^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \qquad (3.27)$$

Then:

$$\begin{aligned}
A(s_t) &= \exp\left[-\frac{1}{2\sigma^2}\left(s_t^2 - 2s_t\left(\frac{h_{1t}}{\sigma_1^2} + \frac{h_{2t}}{\sigma_2^2}\right)\sigma^2\right)\right] \qquad (3.28) \\
&= \exp\left[-\frac{1}{2\sigma^2}\left(\left(s_t - \left(\frac{h_{1t}}{\sigma_1^2} + \frac{h_{2t}}{\sigma_2^2}\right)\sigma^2\right)^2 - \left(\frac{h_{1t}}{\sigma_1^2} + \frac{h_{2t}}{\sigma_2^2}\right)^2\sigma^4\right)\right] \qquad (3.29) \\
&= \exp\left[-\frac{1}{2\sigma^2}\left(s_t - \left(\frac{h_{1t}}{\sigma_1^2} + \frac{h_{2t}}{\sigma_2^2}\right)\sigma^2\right)^2 + \frac{1}{2}\sigma^2\left(\frac{h_{1t}}{\sigma_1^2} + \frac{h_{2t}}{\sigma_2^2}\right)^2\right]. \qquad (3.30)
\end{aligned}$$

Finally:

$$\begin{aligned}\int_{-\infty}^{+\infty} A\left(s_t\right) ds_t &= \exp\left[\frac{1}{2}\sigma^2\left(\frac{h_{1t}}{\sigma_1^2}+\frac{h_{2t}}{\sigma_2^2}\right)^2\right]\sigma\sqrt{2\pi} \\ &\quad \times \underbrace{\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left[-\frac{1}{2\sigma^2}\left(s_t-\left(\frac{h_{1t}}{\sigma_1^2}+\frac{h_{2t}}{\sigma_2^2}\right)\sigma^2\right)^2\right]ds_t}_{=1}. \quad (3.31)\end{aligned}$$

Since the integral of the Gaussian density with mean $\mu_t \equiv \left(\frac{h_{1t}}{\sigma_1^2}+\frac{h_{2t}}{\sigma_2^2}\right)\sigma^2$ and variance σ^2 is equal to 1, we obtain:

$$\int_{-\infty}^{+\infty} A\left(s_t\right) ds_t = \exp\left[\frac{\sigma^2}{2}\left(\frac{h_{1t}}{\sigma_1^2}+\frac{h_{2t}}{\sigma_2^2}\right)^2\right]\sigma\sqrt{2\pi}. \quad (3.32)$$

Therefore:

$$\Lambda_{ML}^{G} = \max_{\alpha,\sigma_1,\sigma_2}\prod_{t=1}^{T}\frac{\sigma\sqrt{2\pi}}{\sqrt{2\pi}\alpha}\frac{\overline{\sigma}_1\overline{\sigma}_2}{\sigma_1\sigma_2}\exp\left[-\frac{h_{1t}^2}{2\sigma_1^2}-\frac{h_{2t}^2}{2\sigma_2^2}+1\right]\exp\left[\frac{\sigma^2}{2}\left(\frac{h_{1t}}{\sigma_1^2}+\frac{h_{2t}}{\sigma_2^2}\right)^2\right] \quad (3.33)$$

$$\begin{aligned}&= \max_{\alpha,\sigma_1,\sigma_2}\prod_{t=1}^{T}\frac{\overline{\sigma}_1\overline{\sigma}_2}{\sqrt{\sigma_1^2\sigma_2^2+\sigma_1^2\alpha^2+\sigma_2^2\alpha^2}} \\ &\quad \times\exp\left[-\frac{h_{1t}^2}{2\sigma_1^2}-\frac{h_{2t}^2}{2\sigma_2^2}+1+\frac{\left(\frac{h_{1t}}{\sigma_1^2}+\frac{h_{2t}}{\sigma_2^2}\right)^2}{2\left(\frac{1}{\alpha^2}+\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)}\right] \quad (3.34)\end{aligned}$$

$$\begin{aligned}&= \max_{\alpha,\sigma_1,\sigma_2}\left\{\left(\frac{\overline{\sigma}_1\overline{\sigma}_2}{\sqrt{\sigma_1^2\sigma_2^2+\sigma_1^2\alpha^2+\sigma_2^2\alpha^2}}\right)^T\right. \\ &\quad \left.\times\exp\left[-\sum_{t=1}^{T}\left(\frac{h_{1t}^2}{2\sigma_1^2}+\frac{h_{2t}^2}{2\sigma_2^2}\right)+T+\frac{1}{2\left(\frac{1}{\alpha^2}+\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)}\sum_{t=1}^{T}\left(\frac{h_{1t}}{\sigma_1^2}+\frac{h_{2t}}{\sigma_2^2}\right)^2\right]\right\} \quad (3.35)\end{aligned}$$

We finally have:

$$\Lambda_{ML}^{G} = \max_{\alpha,\sigma_1,\sigma_2} \left\{ \left(\frac{\overline{\sigma}_1 \overline{\sigma}_2}{\sqrt{\sigma_1^2\sigma_2^2 + \sigma_1^2\alpha^2 + \sigma_2^2\alpha^2}} \right)^T \times \exp\left[-\frac{T\overline{\sigma}_1^2}{2\sigma_1^2} - \frac{T\overline{\sigma}_2^2}{2\sigma_1^2} + T + \frac{T}{2\left(\frac{1}{\alpha^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)} \left(\frac{\overline{\sigma}_1^2}{\sigma_1^4} + \frac{\overline{\sigma}_2^2}{\sigma_2^4} + \frac{2\overline{\alpha}^2}{\sigma_1^2\sigma_2^2} \right) \right] \right\} \quad (3.36)$$

where:

$$\overline{\alpha}^2 = \frac{1}{T} \sum_{t=1}^{T} h_{1t} h_{2t}. \quad (3.37)$$

In the end, we obtain:

$$\Lambda_{ML}^{G} = \max_{\alpha,\sigma_1,\sigma_2 \geq 0} \left\{ \frac{\overline{\sigma}_1 \overline{\sigma}_2}{\sqrt{\sigma_1^2\sigma_2^2 + \sigma_1^2\alpha^2 + \sigma_2^2\alpha^2}} \exp\left[\frac{\frac{\overline{\sigma}_1^2}{\sigma_1^4} + \frac{\overline{\sigma}_2^2}{\sigma_2^4} + \frac{2\overline{\alpha}^2}{\sigma_1^2\sigma_2^2}}{2\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{1}{\alpha^2}\right)} - \frac{\overline{\sigma}_1^2}{2\sigma_1^2} - \frac{\overline{\sigma}_2^2}{2\sigma_2^2} + 1 \right] \right\}^T, \quad (3.38)$$

which is equation (3.9) in [62]. One can show that the maximum is reached for:

$$\alpha^2 = \widehat{\alpha}^2 \equiv \left(\overline{\alpha}^2\right)^+ \quad (3.39)$$

$$\sigma_i^2 = \widehat{\sigma}_i^2 \equiv \left(\overline{\sigma}_i^2 - \widehat{\alpha}^2\right)^+ \quad (3.40)$$

where $(x)^+ = x$ if $x > 0$ and $(x)^+ = 0$ otherwise. This positivity restriction arises because of the positivity constraints on α, σ_1 and σ_2 in the maximization procedure. Maximum-likelihood estimators are unbiased but have no optimum properties for finite samples, in the sense that (when evaluated on finite samples) other estimators may have greater concentration around the true parameter value. However, ML estimators do possess a number of attractive limiting properties [114]: As the sample size increases to infinity, sequences of maximum likelihood estimators are known to be (1) consistent, meaning that the sequence of ML estimators converges in probability to the true unknown population value being estimated and (2) asymptotically normal, meaning that as the sample size increases the distribution of the ML estimator tends to the Gaussian distribution with a

mean equal to the true population value and a covariance matrix equal to the inverse of the Fisher information matrix:

$$\mathcal{I}_{ij} \equiv -\mathbb{E}\left[\frac{\partial^2 \log \Lambda_{ML}}{\partial\theta_i \partial\theta_j}\right], \tag{3.41}$$

where $\theta = (\theta_i)_{i=1,\ldots,N}$ is the vector of unknown parameters of size N. In the Gaussian case, $\theta = (\alpha, \sigma_1, \sigma_2)$ and $N = 3$. Beside, the ML estimators are asymptotically efficient, meaning that no consistent estimator has lower asymptotic mean squared error than the ML estimator, an asymptotic mean squared error given by the Cramér-Rao lower bound

$$\mathbb{V}ar\left(\widehat{\theta}_i^{ML}\right) \geq \mathcal{I}_{ii}^{-1}, \tag{3.42}$$

where $\mathcal{I}^{-1}$ where is the reciprocal of the norm of the Fisher information matrix. Shifting from the estimation problem to the detection problem, the corresponding detection statistic is:

$$\Lambda_{ML}^{G} = \left(1 - \frac{\widehat{\alpha}^4}{\overline{\sigma}_1^2 \overline{\sigma}_2^2}\right)^{-T/2}. \tag{3.43}$$

This detection statistic is shown to be optimal in the sense that it yields the smallest probability of mistakenly concluding a signal is absent (probability of a false dismissal, or pfd) after choosing the threshold that fixes the probability for mistakenly concluding a signal is present (probability of a false alarm, or pfa). Concretely, to determine whether or not the data h contains some desired signal, one then compares the value of this detection statistic to some threshold value $\overline{\Lambda}$. If Λ_{ML}^{G} is greater than the threshold value $\overline{\Lambda}$, one concludes that a signal is present and otherwise one concludes that no signal is present. The so-called cross-correlation statistic Λ_{cc}^{G} can then be obtained from the ML statistic via a monotonic transformation chosen so as to remove the time-dependency [62]:

$$\Lambda_{cc}^{G} = \sqrt{1 - \left(\Lambda_{ML}^{G}\right)^{-2/T}} = \frac{\widehat{\alpha}^2}{\overline{\sigma}_1 \overline{\sigma}_2}. \tag{3.44}$$

We may also characterize the "strength" of a stochastic background in terms of the

signal-to-noise ratio $\frac{S}{N}$ of the cross-correlation statistic Λ_{cc}^{G}, which is simply given by [62]:

$$\frac{S}{N} = \frac{\alpha^2\sqrt{T}}{\sigma_1\sigma_2}. \tag{3.45}$$

Note that the signal-to-noise ratio grows with the squared-root of time so allowing for longer observations naturally implies the possibility to detect fainter signals.

Extending the Analysis to Detectors that are not Colocated and/or Coaligned

While the analysis discussed so far has been based upon the explicit assumption that the two detectors were colocated and coaligned, it can be extended to the general case of detectors that are not located in the same place and/or have different arm orientations. For this extension, it is more convenient to switch from the time-series domain to the frequency domain. A purely stochastic background of GWs is expected to be isotropic, stationary and unpolarized. Its main properties are thus best described in terms of frequency spectrum, which can be expressed in a number of equivalent forms: (1) in the form of a (dimensionless) energy density per unit logarithmic interval of frequency $\Omega_{GW}(f)$, (2) in the form of the spectral density of the average of the Fourier component of the metric $S_s(f) = s_s^2(f)$, where $s_s^2(f)$ has dimension $\mathrm{Hz}^{-\frac{1}{2}}$, or (3) in the form of a (dimensionless) characteristic amplitude of the stochastic background $S_{s,c}(f)$. This latter quantity is typically compared to the detector sensitivity curve $S_{n,c}(f)$, also a dimensionless quantity. In what follows we present a number of relationships relating these variables, and also discuss their use in SGWB signal detection. This discussion is largely inspired by [53], as well as additional references that are specified below.

Characterization of SGWB Signals in the Frequency Domain We first express the frequency spectrum of the stochastic background in terms of the GW *energy density spectrum* $\Omega_{GW}(f)$, defined as the fractional contribution of the energy density in gravitational waves to the total energy density needed to close the universe, which, from

equation 2.109, can be defined as:

$$\Omega_{GW}(f) = \frac{1}{\rho_{crit}} \frac{d\rho_{GW}(f)}{d\ln f}, \tag{3.46}$$

where $\rho_{crit} = \frac{3c^2 H_0^2}{8\pi G}$ is the critical density required to close the universe (see equation 2.98), and where $d\rho_{GW}(f)$ is the energy density between frequencies f and $f + df$.

SGWB signals can also be characterized in terms of their *power spectral density* S_s (here the subscript s stands for signal), which is defined as:

$$S_s(f) = \widetilde{s}^*(f)\,\widetilde{s}(f), \tag{3.47}$$

where $\widetilde{s}^*$ is the complex conjugate of $\widetilde{s}$ and where $\widetilde{s}(f)$ is the Fourier transform of the signal:

$$\widetilde{s}(f) = \int_{-\infty}^{+\infty} e^{-2\pi i f t} s_t dt. \tag{3.48}$$

The power spectral density enters the definition of the average squared fluctuations, or variance, in the signal strain intensity through:

$$\mathbb{V}ar(S) = \alpha^2 = \int_0^{\infty} S_s(f)\, df, \tag{3.49}$$

Note that the power spectral density is defined in 3.49 with a one-sided convention (when the power spectral density is defined with a two-sided convention, the above integral is taken from $-\infty$ to $+\infty$). Similar quantities can also be defined for the noise distribution and also for the measurement output h, and we may in particular introduce the noise power spectral density S_n:

$$S_n(f) = \widetilde{n}^*(f)\,\widetilde{n}(f), \tag{3.50}$$

where $\widetilde{n}(f)$ is the Fourier transform of the noise distribution:

$$\widetilde{n}(f) = \int_{-\infty}^{+\infty} e^{-2\pi i f t} n_t dt. \tag{3.51}$$

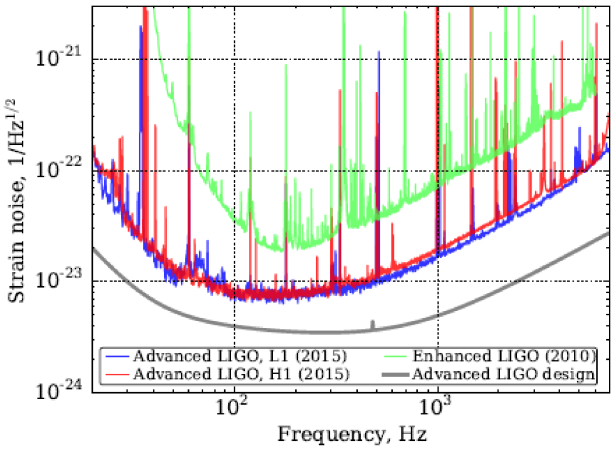

Figure 3.3: The strain sensitivity for the LIGO Livingston detector (L1) and the LIGO Hanford detector (H1) during O1. Also shown is the noise level for the Advanced LIGO design (gray curve) and the sensitivity during the final data collection run (S6) of the initial detectors. This picture is borrowed from [123], where it appears as Figure 2.

Other important quantities are the signal amplitude spectral density defined as $s_s(f) = \sqrt{S_s(f)}$ which has dimension $\mathrm{Hz}^{-\frac{1}{2}}$ and the so-called *characteristic* strain, a dimensionless quantity denoted by $S_{s,c}(f)$ and defined as:

$$S_{s,c}(f) = \sqrt{f S_s(f)}. \tag{3.52}$$

Intuitively, the characteristic strain represents the root mean squared signal in a frequency interval of width $\Delta f = f$ centered at frequency f. The characteristic strain of a signal as a function of the frequency is then compared to the detector sensitivity curve $S_{n,c}(f) = \sqrt{f S_n(f)}$, which is shown in Figure 3.3 for the Advanced LIGO detector (picture borrowed from [123]).

We assume that the random metric perturbations associated with the stochastic background generated from the aggregation of unresolved sources of astrophysical or cosmological origins have no preferred directions so the background can be assumed to be unpolarized and isotropic. As a result, it can be decomposed into gravitational plane waves coming from every direction and polarization. In this situation with unpolarized

and isotropic SGWBs, the gravitational-wave power spectral density can be related to the energy density spectrum via the following relationship, using again a one-sided convention for the power sectral density ([169]):

$$\Omega_{GW}(f) = \frac{4\pi^2}{3H_0^2} f^3 S_s(f). \tag{3.53}$$

In terms of the characteristic strain $S_{s,c}(f) = \sqrt{f S_s(f)}$, we obtain:

$$\Omega_{GW}(f) = \frac{4\pi^2}{3H_0^2} f^2 S_{s,c}^2(f). \tag{3.54}$$

We sometimes assume a power-law frequency dependency for the spectral density of the signal. For a background arising from binary coalescence, we have for example ([169]):

$$\Omega_{GW}(f) \propto f^{\frac{2}{3}}. \tag{3.55}$$

GWSB Signal Processing with Detectors that are not Colocated and Coaligned

For a pair of detectors, the SGWB signals are identical if the detectors are colocated and coaligned, which implies identical cross-spectral densities. When the two detectors are not colocated and/or not coaligned they will have a different response to the SGWB signal, and we expect a reduction in sensitivity due to the violation of the co-location and co-alignment assumptions. The violation of co-alignment manifests itself in a non-parallel alignment of the detector arms. The violation of co-location, on the other hand, manifests itself in a time delay between the observation of the same signal by the two detectors. For example, the inter-site propagation time between the two LIGO detectors located respectively in Hanford, Washington and in Livingston, Louisiana, has been 7 ms for GW150914. This reduction of sensitivity effect is captured by an *overlap reduction function*, which will enter the expression for the first moment of the cross-correlation statistic, as will be seen in equation 3.65.

We first consider the general expression for a cross-correlation of the output of two

detectors operating for a time T (e.g., one year):

$$Y = \int_{-T/2}^{T/2} dt \int_{-T/2}^{T/2} dt' h_1(t)\, h_2(t')\, Q(t,t'), \tag{3.56}$$

where $Q(t-t')$ is a general (real) filter function. When both the SGWB signal and the detector noise are assumed to be stationary, the optimal filter function Q can only depend upon the time-interval $t'-t$ so we can restrict the search of a filter function to the functions of the form $Q(t,t') = Q(t-t')$. Switching to the frequency domain, we write the cross-correlation product [17]:

$$Y = \int_{-T/2}^{T/2} df \int_{-T/2}^{T/2} df' \delta_T(f-f')\, \widetilde{h}_1^*(f)\, \widetilde{h}_2^*(f')\, \widetilde{Q}(f'), \tag{3.57}$$

where $\widetilde{h}_i^*(f)$ is the Fourier transform of the measurement output $h_i(t) = n_i(t) + s(t)$ of detector i (here $i = 1, 2$). The function $\delta_T(f-f')$ is the finite time approximation of the Dirac delta function, which converges to the ordinary delta function in the limit of a large T:

$$\delta_T(f) = \int_{-T/2}^{+T/2} df e^{-i2\pi ft} \underset{T\to\infty}{\to} \delta(f), \tag{3.58}$$

with $\delta_T(0) = T$. Finally $\widetilde{Q}(f)$ is the Fourier transform of the suitably chosen filter function, which we discuss now.

The optimal choice of a filter function depends on the optimization objective. In the context of a GWSB detection, the natural approach, as recalled in our discussion of the time-series analysis, is to minimize the probability of mistakenly concluding a signal is absent (probability of a false dismissal, or pfd) after choosing the threshold that fixes the probability for mistakenly concluding a signal is present (probability of a false alarm, or pfa). If we assume that both the signal and the noise are Gaussian, one can show [17] that the optimization criterion $\min pfd$ subject to a fixed pfa is equivalent to maximizing the signal to noise ratio:

$$\frac{S}{N} = \frac{\mathbb{E}(Y)}{\sqrt{\mathbb{V}ar(Y)}}. \tag{3.59}$$

Note that the situation is different in a non-Gaussian context, where the maximization of a signal to noise ratio involving the first two moments of the cross-correlation statistic will a priori not be equivalent to minimizing the probability of a false alarm for a given probability of a false dismissal. We shall revisit this question of optimal detection statistic in the presence of a deviation from the Gaussian assumption in chapter 5, section 5.3.3.

Maintaining for the moment a sole focus on the first two moments of the cross-correlation statistic, and assuming independent noise distributions with a zero mean, we have:

$$\mathbb{E}\left(Y\right) = \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} df' \delta_T \left(f - f'\right) \mathbb{E}\left[\widetilde{s}_1^*\left(f\right) \widetilde{s}_2^*\left(f'\right)\right] \widetilde{Q}\left(f'\right). \tag{3.60}$$

Using a plane wave expansion of the SGWB signal and taking the ensemble average, it can be shown [17] that:

$$\mathbb{E}\left[\widetilde{s}_1^*\left(f\right) \widetilde{s}_2^*\left(f'\right)\right] = \frac{3H_0^2}{20\pi^2} \left|f\right|^{-3} \Omega_{GW}\left(\left|f\right|\right) \gamma\left(\left|f\right|\right), \tag{3.61}$$

where $\gamma\left(f\right)$ is a dimensionless function of the frequency known as the *overlap reduction function* [79] that captures the impact of deviations from the ideal situation with co-located and co-aligned detectors:

$$\begin{aligned} \gamma\left(f\right) &= \frac{5}{8\pi} \int_{S^2} d\widehat{\Omega} e^{i2\pi f \widehat{\Omega}.\Delta\overrightarrow{x}/c} F_1^+\left(\widehat{\Omega}\right) F_2^+\left(\widehat{\Omega}\right) \\ &\quad + \frac{5}{8\pi} \int_{S^2} d\widehat{\Omega} e^{i2\pi f \widehat{\Omega}.\Delta\overrightarrow{x}/c} F_1^\times\left(\widehat{\Omega}\right) F_2^\times\left(\widehat{\Omega}\right), \end{aligned} \tag{3.62}$$

where $\widehat{\Omega}$ is a unit vector specifying the direction on the two-sphere, $\overrightarrow{x} = \overrightarrow{x}_1 - \overrightarrow{x}_2$ is the separation vector between the central stations of the two detectors, and where

$$F_i^+\left(\widehat{\Omega}\right) = e_{ab}^+\left(\widehat{\Omega}\right) d_i^{ab} \tag{3.63}$$

$$F_i^\times\left(\widehat{\Omega}\right) = e_{ab}^\times\left(\widehat{\Omega}\right) d_i^{ab} \tag{3.64}$$

are the responses of the i^{th} $(i = 1, 2)$ detector to the $+$ and $\times$ polarizations [17]. Here

d_i^{ab} is a symmetric tracefree tensor that specifies the orientation of the two arms of the i^{th} $(i = 1, 2)$ detector. Given the normalization constant $\frac{5}{8\pi}$, we have $\gamma(f) = 1$ for all f for co-located and co-aligned detectors, and $\gamma(f) < 1$ when the detectors are separated and/or not aligned.

Plugging 3.61 into 3.60, we obtain:

$$\mathbb{E}(Y) = \frac{3H_0^2}{20\pi^2} T \int_{-\infty}^{\infty} df\, |f|^{-3}\, \Omega_{GW}(|f|)\, \gamma(|f|)\, \widetilde{Q}(f)\,. \tag{3.65}$$

It can also be shown [17] that when the noise is much larger than the signal, the variance of the cross-correlation statistic can be approximated by:

$$\mathbb{V}ar(Y) \approx \frac{T}{4} \int_{-\infty}^{\infty} df S_{n_1}(|f|)\, S_{n_2}(|f|) \left|\widetilde{Q}(f)\right|^2, \tag{3.66}$$

where

$$\left|\widetilde{Q}(f)\right|^2 = \widetilde{Q}^*(f)\, \widetilde{Q}(f)\,, \tag{3.67}$$

and where $S_{n_1}(f)$ and $S_{n_2}(f)$ are the power spectral densities of the noise in the two detectors.[4] If we now define the (positive definite) scalar product of two arbitrary complex functions:

$$(A, B) \equiv \int_{-\infty}^{\infty} df A^*(f)\, B(f)\, S_{n_1}(|f|)\, S_{n_2}(|f|)\,, \tag{3.68}$$

the expressions for the mean 3.65 and variance 3.66 of the cross-correlation statistic can be rewritten as:

$$\mathbb{E}(Y) = \frac{3H_0^2}{20\pi^2} T \left(\widetilde{Q}, \frac{|f|^{-3}\, \Omega_{GW}(|f|)\, \gamma(|f|)}{S_{n_1}(|f|)\, S_{n_2}(|f|)}\right), \tag{3.69}$$

$$\mathbb{V}ar(Y) \approx \frac{T}{4}\left(\widetilde{Q}, \widetilde{Q}\right). \tag{3.70}$$

Going back to the optimal choice of the filter function so as to maximize the signal to

[4] Note that $S_{n_i}(|f|) = 2S_{ni}(f)$ is the one-sided power spectrum of noise on detector i.

noise ratio $\frac{S}{N}$ we obtain:

$$\left(\frac{S}{N}\right)^2 = \frac{\left[\mathbb{E}\left(Y\right)\right]^2}{\mathbb{V}ar\left(Y\right)} \approx \left(\frac{3H_0^2}{20\pi^2}\right)^2 4T \frac{\left(\widetilde{Q}, \frac{|f|^{-3}\Omega_{GW}(|f|)\gamma(|f|)}{S_{n_1}(|f|)S_{n_2}(|f|)}\right)^2}{\left(\widetilde{Q}, \widetilde{Q}\right)}. \tag{3.71}$$

We now need to find the filter function $\widetilde{Q}$ that maximizes $\left(\frac{S}{N}\right)^2$, a quantity that can be written as $\frac{\left(\widetilde{Q},A\right)^2}{\left(\widetilde{Q},\widetilde{Q}\right)}$ with

$$A = \frac{|f|^{-3}\,\Omega_{GW}\left(|f|\right)\gamma\left(|f|\right)}{S_{n_1}\left(|f|\right)S_{n_2}\left(|f|\right)}. \tag{3.72}$$

This ratio is proportional to the squared cosine of the angle between the vectors $\widetilde{Q}$ and A, and is therefore maximized by taking $\widetilde{Q}$ pointing in the same direction as A, that is by taking:

$$\widetilde{Q}\left(f\right) = \lambda \frac{\gamma\left(|f|\right)\Omega_{GW}\left(|f|\right)}{|f|^3\, S_{n_1}\left(|f|\right)S_{n_2}\left(|f|\right)} \tag{3.73}$$

for some normalization constant λ. Note that the optimal filter function 3.73 depends on the SGWB frequency spectrum $\Omega_{GW}\left(f\right)$, which is not known a priori. In practice, and as mentioned before, the standard approach consists in assuming a power-law frequency dependency for the spectral density of the signal:

$$\Omega_{GW}\left(f\right) = \Omega_\alpha f^\alpha, \tag{3.74}$$

where Ω_α is a constant. Note that $\alpha = 0$ corresponds to the case of a power spectrum that would be a constant $\Omega_{GW}\left(f\right) = \Omega_\alpha$. A set of optimal filter functions indexed by α are then constructed

$$\widetilde{Q}_\alpha\left(f\right) = \lambda_\alpha \frac{\gamma\left(|f|\right)|f|^\alpha}{|f|^3\, S_{n_1}\left(|f|\right)S_{n_2}\left(|f|\right)}, \tag{3.75}$$

where the normalizing constant λ_α is chosen so that $\mathbb{E}\left(Y\right) = \Omega_\alpha T$. With this normalization choice, the optimal filter function only depends on the exponent α, the overlap reduction function and the power spectrum for the noise on each one of the two detectors.

Finally, the signal-to-noise ratio with the optimal filter is given by [53]:

$$\frac{S}{N} = \frac{\mathbb{E}\left(Y\right)}{\sqrt{\mathbb{V}ar\left(Y\right)}} = \lambda \frac{3H_0^2}{10\pi^2}\sqrt{T}\sqrt{2\int_0^\infty \frac{\gamma^2\left(f\right)}{S_{n_1}\left(f\right)S_{n_2}\left(f\right)}\frac{\Omega_{GW}^2\left(f\right)}{f^6}df}. \tag{3.76}$$

We note again that the signal-to-noise ratio grows with the squared-root of time, so longer observation times allow for the detection of fainter signals. We may also compute the minimum detectable amplitude by a pair of detectors [17]:

$$\Omega_{\min} = \frac{4\pi^2}{3H_0^2\sqrt{T}}\left[\mathrm{erfc}^{-1}\left(2\alpha\right) - \mathrm{erfc}^{-1}\left(2\left(1-\beta\right)\right)\right]\sqrt{\int_0^\infty \frac{\gamma^2\left(f\right)}{f^6 S_{n_1}\left(f\right)S_{n_2}\left(f\right)}df}, \tag{3.77}$$

where for $\alpha = pfa$ is the probability of a false alarm, $\beta = pfd$ is the probability of a false dismissal, and erfc is the complementary error function:

$$\mathrm{erfc}\left(x\right) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}dt. \tag{3.78}$$

The minimum detectable amplitude is a very useful quantity which is often used as a measure of detector sensitivity.

3.2.2 Bayesian Approach to SGWB Data Analysis

The key difference between the frequentist and Bayesian approaches to statistics can be summarized as follows. In frequentist statistics, probabilities are fundamentally related to frequencies of events and the only input to draw statistical conclusions from is the sampled data. In Bayesian statistics, probabilities are fundamentally related to some prior knowledge, if any, about a particular hypothesis or paramter value, and the inputs to draw statistical conclusions from includes not only the sampled data, but also some prior subjective view formed by the statistician before even looking at the sample data. In Bayesian statistics, Bayes' law (equation 3.79) is used to update the prior for a hypothesis (about the presence of a GW signal for the detection question, or about its parameter values for the estimation question) as a function of the sampled data, so as to turn the

prior into a qualified posterior view (about the presence of a signal or about its parameter values). Let us for example consider the estimation question, where the problem is summarized in terms of the need to estimate some possibly multi-dimensional parameter Θ given some data set containing the measurement output $\mathcal{H}$. According to Bayes law, the posterior probability distribution that the parameter Θ takes on a given value θ given the data $\mathcal{H}$ is given by:[5]

$$\underbrace{\Pr\left(\Theta = \theta | \, \mathcal{H}\right)}_{\text{posterior probability}} = \frac{\underbrace{\Pr\left(\mathcal{H} | \, \theta\right)}_{\text{likelihood}} \times \underbrace{\Pr\left(\theta\right)}_{\text{prior probability}}}{\underbrace{\Pr\left(\mathcal{H}\right)}_{\text{marginalized likelihood}}}, \tag{3.79}$$

where the marginalized likelihood $\Pr\left(\mathcal{H}\right)$ is given by:

$$\Pr\left(\mathcal{H}\right) = \int \Pr\left(\mathcal{H} | \, \theta\right) \Pr\left(\theta\right) d\theta, \tag{3.80}$$

and where $\Pr\left(\mathcal{H} | \, \theta\right)$ can be interpreted as the probability of observing the data for a given value of the parameter. Extending the analysis for all possible values of the parameter, we obtain:

$$\underbrace{\Pr\left(\Theta | \, \mathcal{H}\right)}_{\text{posterior probability distribution}} = \frac{\underbrace{\Pr\left(\mathcal{H} | \, \Theta\right)}_{\text{likelihood}} \times \underbrace{\Pr\left(\Theta\right)}_{\text{prior probability distribution}}}{\underbrace{\Pr\left(\mathcal{H}\right)}_{\text{marginalized likelihood}}}. \tag{3.81}$$

Given a posterior distribution, a Bayesian confidence interval (often called a credible interval given the Bayesian interpretation of probability as state of knowledge about an event) can be defined similarly to confidence intervals in the frequentist approach. One can also use the Bayesian framework to generate a point estimate for the unknown parameter. To do so, let us assign a cost function $C\left(\Theta', \Theta\right)$ of estimating the true value of Θ as Θ'. We then associate with an estimator $\widehat{\Theta}$ the cost function for the estimator

[5] Note that we use the notation $\mathbf{H}$ to denote an hyothesis, and the notation $\mathcal{H}$ to denote measurement output.

averaged over all realizations of the data in the sample for each value θ of the parameter Θ [102]:

$$R_\theta\left(\widehat{\Theta}\right) \equiv \mathbb{E}_\theta\left[C\left(\widehat{\Theta},\theta\right)\right] = \int_{\mathcal{H}} C\left(\widehat{\Theta}\left(h\right),\theta\right)\Pr\left(h,\Theta=\theta\right)dh. \tag{3.82}$$

We then introduce the Bayes estimator as the estimator that minimizes the average risk defined as the average of the cost function over all parameter values θ:

$$r\left(\widehat{\Theta}\right) \equiv \mathbb{E}\left[R_\theta\left(\widehat{\Theta}\right)\right] = \int_{\Theta}\int_{\mathcal{H}} C\left(\widehat{\Theta}\left(h\right),\theta\right)\Pr\left(h,\theta\right)\Pr\left(\theta\right)dhd\theta. \tag{3.83}$$

It is easy to show that for a commonly used cost function:

$$C\left(\Theta',\Theta\right) = \left\|\Theta'-\Theta\right\|^2, \tag{3.84}$$

the Bayesian estimator is simply the mean of the posterior density, i.e., the conditional mean of the parameter Θ given data $\mathcal{H}$:

$$\widehat{\Theta}_{Bayesian} = \mathbb{E}\left[\Theta|\,\mathcal{H}\right] = \int_{\Theta}\theta\Pr\left(\Theta=\theta|\,\mathcal{H}\right)d\theta, \tag{3.85}$$

a quantity which depends of course upon the assumed prior distribution $\Pr\left(\Theta\right)$.

Turning now to the detection question, and denoting by $\mathbf{H}_1$ the null hypothesis that a signal is present in the data $\mathcal{H}$, and by $\mathbf{H}_0$ the alternative hypothesis that the signal is absent, we have:

$$\underbrace{\Pr\left(\mathbf{H}_1|\,\mathcal{H}\right)}_{\text{posterior probability}} = \frac{\underbrace{\Pr\left(\mathcal{H}|\,\mathbf{H}_1\right)}_{\text{likelihood}} \times \underbrace{\Pr\left(\mathbf{H}_1\right)}_{\text{prior probability}}}{\underbrace{\Pr\left(\mathcal{H}\right)}_{\text{marginalized likelihood}}}, \tag{3.86}$$

where the normalizing factor, known as the marginalized likelihood $\Pr\left(\mathcal{H}\right)$, is given by:

$$\Pr\left(\mathcal{H}\right) = \Pr\left(\mathcal{H}|\,\mathbf{H}_1\right)\Pr\left(\mathbf{H}_1\right) + \Pr\left(\mathcal{H}|\,\mathbf{H}_0\right)\Pr\left(\mathbf{H}_0\right). \tag{3.87}$$

Thus, an optimal detection statistic is uniquely determined in the Bayesian framework as

a threshold on the posterior probability distribution; however, that statistic does depend on choices of prior probability distributions.

The Bayesian approach can also be used to compare two models $\mathcal{M}_i$ and $\mathcal{M}_j$, where the notion of a "model" refers to a set of parametric assumptions for the assumed distribution of the measurement output $\mathcal{H}$. To do so, we first compute the posterior probability for each model:

$$\underbrace{\Pr\left(\mathcal{M}_i|\,\mathcal{H}\right)}_{\text{posterior probability}} = \frac{\underbrace{\Pr\left(\mathcal{H}|\,\mathcal{M}_i\right)}_{\text{likelihood}} \times \underbrace{\Pr\left(\mathcal{M}_i\right)}_{\text{prior probability}}}{\underbrace{\Pr\left(\mathcal{H}\right)}_{\text{marginalized likelihood}}}. \tag{3.88}$$

To compare two models, we then take the ratio of the posterior probabilities for these two models:

$$\frac{\Pr\left(\mathcal{M}_i|\,\mathcal{H}\right)}{\Pr\left(\mathcal{M}_j|\,\mathcal{H}\right)} = \frac{\Pr\left(\mathcal{H}|\,\mathcal{M}_i\right)}{\Pr\left(\mathcal{H}|\,\mathcal{M}_j\right)} \times \frac{\Pr\left(\mathcal{M}_i\right)}{\Pr\left(\mathcal{M}_j\right)}, \tag{3.89}$$

where the common term $\Pr\left(\mathcal{H}\right)$ has cancelled out. Here $\frac{\Pr(\mathcal{M}_i|\mathcal{H})}{\Pr(\mathcal{M}_j|\mathcal{H})}$ is known as the *posterior odd ratio*, which is shown to be given by the product of the *Bayes factor* $\frac{\Pr(\mathcal{H}|\mathcal{M}_i)}{\Pr(\mathcal{H}|\mathcal{M}_j)}$ and the *prior odd ratio* $\frac{\Pr(\mathcal{M}_i)}{\Pr(\mathcal{M}_j)}$. In the common case when there is no a priori reason to favor a particular model (in Bayesian statistics this case is sometimes referred to as the *non informative* prior case), we take $\Pr\left(\mathcal{M}_i\right) = \Pr\left(\mathcal{M}_j\right)$, and we obtain in this case that the posterior odd ratio is simply the Bayes factor:

$$\frac{\Pr\left(\mathcal{M}_i|\,\mathcal{H}\right)}{\Pr\left(\mathcal{M}_j|\,\mathcal{H}\right)} = \frac{\Pr\left(\mathcal{H}|\,\mathcal{M}_i\right)}{\Pr\left(\mathcal{H}|\,\mathcal{M}_j\right)}. \tag{3.90}$$

If we take one of the models, denoted by $\mathcal{M}_0$, to be a benchmark model, then we can compute the Bayes factors, and therefore the posterior odd ratios, of all the other models relative to $\mathcal{M}_0$. Taking the $\mathcal{M}_0$ model to refer to a situation where the signal is absent and some assumption (say Gaussian) is made for the noise distribution, we can form a detection statistic by an analysis of the *posterior odd ratio* with respect to some assumed model versus the noise-only model.

One might wonder if a Bayesian approach with a non informative prior can be related

to the frequentist approach, where no prior are used as inputs. A formal answer to this question has been given in a recent paper [49], where it is shown that the Bayes factor is proportional to the frequentist maximum-likelihood ratio, with a proportionality constant given by the Occam's factor (see 3.94 below), which penalizes a model if its parameter space volume is larger than necessary to fit the data.[6] More precisely, let us remark that the detection statistic used in the frequentist analysis is obtained by maximizing the likelihood ratio for a signal+noise model $\mathcal{M}_1$ to the noise-only model $\mathcal{M}_0$. For example if a Gaussian assumption is maintained for both the signal and noise distributions, the likelihood ratio 3.13 is given by:

$$\Lambda_{ML}\left(\mathcal{H}\right) = \frac{\max\limits_{\alpha,\sigma_1,\sigma_2} \Pr\left(\mathcal{H}|\,\alpha,\sigma_1,\sigma_2,\mathcal{M}_1\right)}{\max\limits_{\sigma_1,\sigma_2} \Pr\left(\mathcal{H}|\,\sigma_1,\sigma_2,\mathcal{M}_0\right)} \tag{3.91}$$

$$= \frac{\max\limits_{\alpha,\sigma_1,\sigma_2} \int \Pr\left(\mathcal{H}|\,\alpha,\sigma_1,\sigma_2,\mathcal{M}_1\right) \Pr\left(\mathcal{S}-\mathcal{H}|\,\alpha,\sigma_1,\sigma_2,\mathcal{M}_1\right)}{\max\limits_{\sigma_1,\sigma_2} \Pr\left(\mathcal{H}|\,\sigma_1,\sigma_2,\mathcal{M}_0\right)} \tag{3.92}$$

where $\mathcal{M}_1$ represents a model with Gaussian signal and Gaussian noise distributions, while $\mathcal{M}_0$ represents a noise-only model with Gaussian noise distributions. The Bayes factor calculation also involves a ratio of two quantities, but instead of maximizing over the parameters, we *marginalize* over the parameters by computing the posterior odd ratio:

$$\frac{\Pr\left(\mathcal{M}_1|\,\mathcal{H}\right)}{\Pr\left(\mathcal{M}_0|\,\mathcal{H}\right)} = \frac{\int\int\int \Pr\left(\mathcal{H}|\,\alpha,\sigma_1,\sigma_2,\mathcal{M}_1\right) \Pr\left(\alpha,\sigma_1,\sigma_2,|\,\mathcal{M}_1\right) d\alpha d\sigma_1 d\sigma_2}{\int\int \Pr\left(\mathcal{H}|\,\sigma_1',\sigma_2',\mathcal{M}_0\right) \Pr\left(\sigma_1',\sigma_2',|\,\mathcal{M}_0\right) d\sigma_1' d\sigma_2'} \tag{3.93}$$

These two expressions can be related to one another as follows [49]:

$$\frac{\Pr\left(\mathcal{M}_1|\,\mathcal{H}\right)}{\Pr\left(\mathcal{M}_0|\,\mathcal{H}\right)} \simeq \Lambda_{ML}\left(\mathcal{H},\widehat{\alpha},\widehat{\sigma}_1,\widehat{\sigma}_2\right) \frac{\Delta V_1/V_1}{\Delta V_0/V_0} \tag{3.94}$$

where $\Lambda_{ML}\left(\mathcal{H},\widehat{\alpha},\widehat{\sigma}_1,\widehat{\sigma}_2\right)$ is the value of the likelihood function for the parameters $\widehat{\alpha}$, $\widehat{\sigma}_1$, $\widehat{\sigma}_2$ given in 3.39 and 3.40 for which this function reaches its maximum value. Here ΔV_i

[6] See also [28] for a discussion about the use of the ratio of the likelihood of observed data to contain a signal to the likelihood of it being a noise fluctuation to provide optimal ranking for the candidate signal events found in an experiment under both the Bayesian and frequentist statistical approaches.

for $i = 0, 1$ is the characteristic width of the likelihood function around its maximum and V_i is the total parameter space volume for the model parameters. The ratio $\Delta V_i/V_i$ can be thought of as an Occam's factor, which penalizes a model $\mathcal{M}_i$ if its parameter space volume is larger than needed to fit the data $\mathcal{H}$. For example, in the model $\mathcal{M}_1$ based on a Gaussian assumption for the signal distribution with a standard deviation α in a range of amplitude $[0, \alpha_{\max}]$, and also a Gaussian assumption for the noise distributions, assuming a common standard deviation σ, we have that $V_1 = \alpha_{\max}$ and $\Delta V_1 = \frac{\sigma}{\sqrt{T}}$ (see appendix A.1 in [49]).

From a practical standpoint, the frequentist approach is convenient since the likelihood function can be computed analytically, at least under the Gaussian assumption. It is in fact one important contribution of our work (see chapters 4 and 5) to show that it is possible to maintain analytical tractability even in the non-Gaussian case by using a parsimonious approximation for an unknown non-Gaussian density function in terms of a polynomial correction with respect to the Gaussian density function. The Bayesian inference approach is also very simple in principle: one applies Bayes' rule to calculate posterior probability distributions given a likelihood function (which specifies the probability of the data given the model and the value of any parameters associated with it) and a prior probability distribution for the model and its parameters. This approach requires, however, a numerical calculation of the marginalized densities in 3.93, which in practice can be computationally extremely intensive, especially for models having a large number of parameters. Fortunately, the recent development of efficient sampling algorithms [76] makes the use of this approach possible, if not easily tractable, and Bayesian analysis has now been integrated in standard methods for GW signal detection and parameter estimation. From the conceptual standpoint, the Bayesian approach is attractive since it can provide a unified treatment of the three types of search methods used for different types of signals (matched filtering for well-modelled deterministic signals such as those from the coalescence from compact binary inspirals, excess power analysis for poorly-modeled deterministic signals, such as those from core-collapse supernovae and cross-correlation

analysis for stochastic signals). For example, one can consider a waveform template used to describe the gravitational wave signal as an informative prior [48], [149].

3.3 Analysis of Non-Gaussian SGWB Distributions

It is typically assumed that the stochastic background of gravitational waves creates metric perturbations that are stationary, isotropic and unpolarized random processes. The GW background is also usually assumed to be Gaussian invoking the central limit theorem, and thus completely characterized by its mean and variance. The key component of the astrophysical SGWB distribution is the number of sources that simultaneously contribute to the total background, which in turn depends upon the duration of each single event that produce an emission of GWs and the interval between two, or more, of these events. Intuitively, if the number of sources is large enough for the time interval between events to be small compared to the duration of a single event, then the waveforms of each single source overlap to create a continuous background which can be assumed to be well approximated by a Gaussian distribution given the Central Limit Theorem (CLT). On the other hand, recent predictions based on population modeling suggest that for many astrophysical models, there may not be enough overlapping sources, resulting in the formation of a non-Gaussian background. In case of short duration burst GW sources in particular, it is possible that their number is sufficiently small for the time interval between events to be long compared to the duration of a single event, in which case the CLT may not strictly apply. In addition to the non-Gaussianity to be expected in the background from astrophysical origin, it has also been shown that the background from cosmic strings could be dominated by a non-Gaussian contribution arising from the closest sources [55, 144]. In what follows, we first present a number of arguments suggesting that deviations from the Gaussian assumption are to be expected for the SGWB of astrophysical origin because of a number of overlapping sources too small for the Central Limit Theorem to apply. We then introduce the Edgeworth expansion, a

formal approximation of a non-Gaussian random variable given by the sum of a finite number of independent and identically distributed sources.

3.3.1 Stylized Analysis of the Distribution of the GW Signal Given by the Superposition of a Random Number of Sources

We now turn to a stylized analysis of the distribution of the sum of a random number of random sources. Our goal is not to provide a formal statistical model of the SGWB, but instead to illustrate in a simplistic setting that noticeable deviations from normality are to be expected given a realistically large number of GW sources of astrophysical origin.

Introducing a Poisson Process Model for the Number of Sources

In this context, we postulate that the occurrence of individual gravitational wave signals take place as a Poisson process, which is a standard model for describing the arrival of events at random times. Denoting the intensity of the Poisson process by λ, the number of signals in $[0, t]$ is a Poisson random variable $N(t)$ with parameter λt:

$$\Pr\left(N\left(t\right)=k\right)=e^{-\lambda t}\frac{\left(\lambda t\right)^{k}}{k!}, \tag{3.95}$$

so that we have for example:

$$\Pr\left(N\left(t\right)=0\right)=e^{-\lambda t}. \tag{3.96}$$

It can easily be shown that:

$$\mathbb{E}\left(N\left(t\right)\right)=\mathbb{V}ar\left(N\left(t\right)\right)=\lambda t. \tag{3.97}$$

The important afore-mentioned distinction between continuous and burst signals can be captured in this setting by formally relating the intensity process λ to the duty cycle DC. If we consider for example K types of astrophysical events with each type $k =$

$1, ..., K$ having a duty cycle DC_k, we may assume that the number of signals of type k in the interval $[0, t]$, denoted by $N_k(t)$, follows a Poisson process with intensity $\lambda_k = DC_k$.

Let us now further assume that individual GW signals are independent and identically-distributed random variable S_i for $i = 1, 2, ...$, which is a legitimate assumption if we restrict our analysis to a given type of astrophysical events, such as compact binary coalescences or core supernova collapses, for example. We next consider the random sum representing the aggregate signal of astrophysical origin in the period $[0, t]$:

$$S_N(t) = \sum_{i=1}^{N(t)} S_i. \tag{3.98}$$

When $N(t)$ has a Poisson distribution, the process $S_N(t)$ is known as a *compound Poisson process*, the first two moments of which can be easily computed. Using conditional expectations and the law of total variance, we have:

$$\mathbb{E}(S_N(t)) = \mathbb{E}\left(\sum_{i=1}^{N(t)} S_i\right) = \mathbb{E}(N(t))\,\mathbb{E}(S) \tag{3.99}$$

$$\mathbb{V}ar(S_N(t)) = \mathbb{V}ar(\sum_{i=1}^{N(t)} S_i) \tag{3.100}$$

$$= \mathbb{E}(N(t))\,\mathbb{V}ar(S) + \mathbb{V}ar(N(t))\,(\mathbb{E}(S))^2 \tag{3.101}$$

$$= \mathbb{E}(N(t))\,\mathbb{E}(S^2) \tag{3.102}$$

If we denote by μ_1 and μ_2, respectively, the first and second non-central moments $\mu_1 = \mathbb{E}(S)$ and $\mu_2 = \mathbb{E}(S^2)$ of the signal S, then the aggregate signal process has the following mean and variance:

$$\mathbb{E}\left(\sum_{i=1}^{N(t)} S_i\right) = \lambda t \mu_1 \tag{3.103}$$

$$\mathbb{V}ar(\sum_{i=1}^{N(t)} S_i) = \lambda t \mu_2 \tag{3.104}$$

In physical applications to GW signals, μ_1 is zero and μ_2 is the variance of the gravitational wave signal. In the limit of t going to infinity, the Central Limit Theorem implies that the aggregate signal $S_N(t)$ will have a Gaussian distribution. For a finite detection time t, on the other hand, one expects to have deviations from the Gaussian distribution (unless each signal is normally distributed), and the resulting non-Gaussian distribution can be approximated by a suitable correction to the Gaussian distribution via the Edgeworth expansion (see next sub-section for more details about the Edgeworth expansion).

Estimating Realistic Values for the Intensity Parameter

So as to assess the expected deviation from the Gaussian distribution for the SGWB of astrophysical origin, we can estimate the realistic values for the intensity parameter λ that can be expected for massive binary systems, which are expected to be the most significant contributors to the SGWB. In particular, one can obtain estimates for the number of binaries generating gravitational-wave signals in a 1 Hz bin as observed by a detector on Earth, taking into account uncertainties in the star formation rate and in the delay time between the formation and coalescence of the binary [180].[7] This quantity is plotted in Figure 1 [180], which is reproduced below as Figure 3.4, for three different types of binaries and for different values of *Zsup*, a quantity defined as the upper limit on the integrated flux (per unit frequency), which is shown to depend on both the emission frequency range and on the maximum redshift considered for the star formation history calculation. From this Figure, it is found that in any 0.1 second-long time-segment corresponding to the lowest observable frequency of 10 Hz there will be on average 10 binary neutron star systems emitting in the Advanced LIGO 10-200 Hz band. It should

[7] A similar analysis is performed in [151], who provides an estimate of the part of the total background of rotating neutron stars that is unresolvable because such signals cannot be distinguished from each other or subtracted from the data of a gravitational wave detector. The resolvability of the background is quantified by the *overlap function*, $N(f, \Delta f, z)$, a generalization of the duty cycle introduced in [150] that gives the expected number of signals, with redshifts smaller than z, that are observed within a frequency bin $[f, f + \Delta f]$, where Δf is the frequency resolution allowed by the detector and the data analysis method. Depending on assumed parameter values, a number $N(f, \Delta f, z) \geq 1$ is expected, corresponding to one or more overlapping signals.

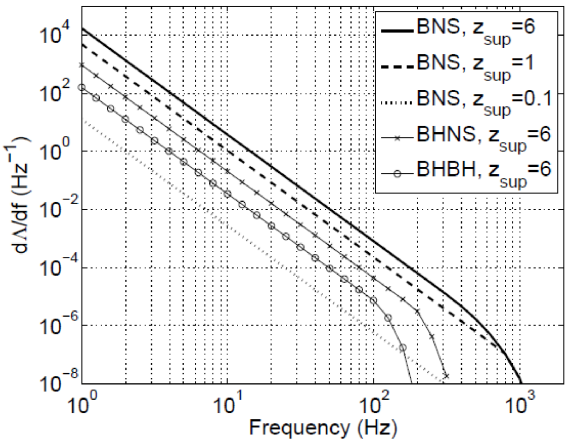

FIG. 1: Number of binaries per 1 Hz frequency bin. For the BNS cases we assume each star to have mass of 1.4 $M_\odot$, local rate of $\lambda = 1\ \mathrm{Mpc}^{-3}\ \mathrm{Myr}^{-1}$, and $t_{\min} = 20$ Myr ($P(t_d) \sim 1/t$). For the BHNS case we assume masses of 1.4 $M_\odot$ and 10 $M_\odot$, the local rate $\lambda = 0.03\ \mathrm{Mpc}^{-3}\ \mathrm{Myr}^{-1}$, and $t_{\min} = 100$ Myr. For the BBH case we assume masses of 10 $M_\odot$, and $t_{\min} = 100$ Myr.

Figure 3.4: This Figure is borrowed from [180]. Note that the notation λ is used in that paper to denote the mass fraction converted into progenitors.

be noted that the duty cycle is somewhat lower for the BBH and BHNS cases.

On the other hand, if we focus on the 100 Hz frequency, we expect a duty cycle around 10^{-3} for binary black holes systems, which would result in strong deviations from the Gaussian situation from the SGWB (see Figures 3.7 and 3.8 below). In fact, we are likely to fall in this case in a non-overlapping regime, with an average number of non-overlapping sources much lower than 1. For the BBH case, the analysis was performed using the assumption of black holes with masses equal to 10 solar masses. In Figure 3.5, we repeat the same analysis for higher mass values consistent with the recent detections. More precisely, we plot $\Lambda(f)$, which is the average number of sources that overlap in a typical frequency band of width f around the frequency f, at a given time. The blue, red and green areas correspond to the subpopulations of BH-BH binary systems with masses similar to the three events GW150914 [9] ($m_1 = 36.2$, $m_2 = 29.1$), LVT151012 [4] ($m_1 = 23, m_2 = 13$), and GW151226 [5] ($m_1 = 14.2, m_2 = 7.5$), including the errors on the event rates. The black curve is for NS-NS binary systems assuming the LIGO

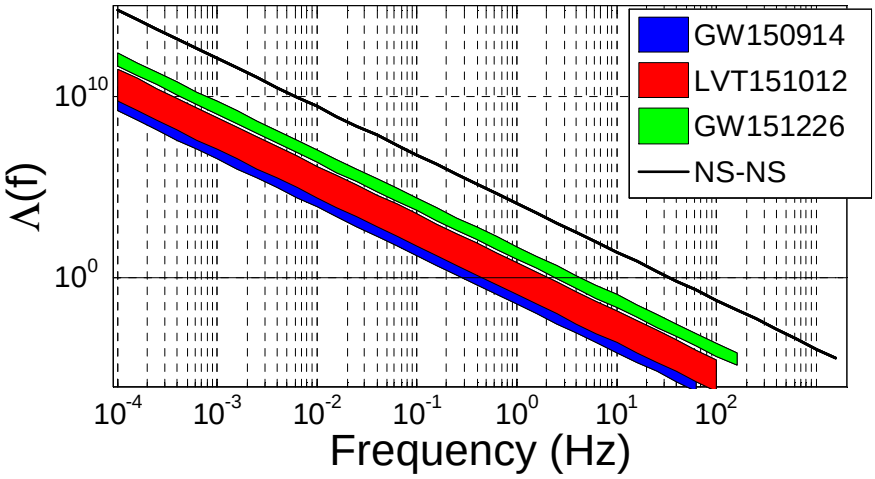

Figure 3.5: This figure shows $\Lambda(f)$, the average number of sources that overlap at a given time in a frequency band given by the frequency resolution, as a function of the frequency f. The blue, red and green areas correspond to the subpopulations of BH-BHs with masses similar to the three events GW150914 [9] ($m_1 = 36.2$, $m_2 = 29.1$), LVT151012 [4] ($m_1 = 23, m_2 = 13$) and GW151226 [5] ($m_1 = 14.2, m_2 = 7.5$), including the errors on the event rates. The black curve is for NS-NS binary systems assuming the rate of 1 Mpc^{-3} Myr^{-1}. The horizontal bar corresponds to $\Lambda = 1$, the threshold above which we obtain a continuous GW background with a signal always present.

realistic rate of 1 Mpc^{-3} Myr^{-1}. The plot is extended down to $f = 10^{-4}$ in order to include the LISA band.[8]

As explained below, we know that the compound Poisson process will converge towards a Gaussian distribution as the intensity process goes to infinity. We now want to provide a quantitative assessment of the actual deviation from normality to be expected for realistic gravitational wave backgrounds so as to be in a situation to analyze what is the fraction of the frequency range that can be assumed to correspond to a Gaussian versus non-Gaussian SGWB signal.

Measuring the Degree of Non Normality as a Function of the Intensity Process

To fix ideas, let us first assume that the stochastic background is defined as the sum of a random number of signals where each signal follows a Delta distribution, meaning that

[8] Since the calculation is stopped at the last stable orbit, we obtain that the end frequency is smaller for larger masses.

the strain intensity is a deterministic fixed quantity, which is normalized to be equal to 1. It is well-known that the Poisson random variable quickly converges towards a Gaussian distribution as the intensity λ increases for a fixed t, or as t increases for a fixed λ (see Figure 3.6, where we take $t \equiv 1$). To prove the convergence, consider the random variable:

$$Y_\lambda(t) = \frac{N(t) - \lambda t}{\sqrt{\lambda t}}. \tag{3.105}$$

Then its characteristic function is:

$$\phi_{Y_\lambda(t)} = \exp\left(-\sqrt{\lambda t} it + \lambda t \left(e^{\frac{it}{\sqrt{\lambda t}}} - 1\right)\right). \tag{3.106}$$

For large λt, we have:

$$e^{\frac{it}{\sqrt{\lambda t}}} - 1 = \frac{it}{\sqrt{\lambda t}} - \frac{t^2}{2\lambda t} + o\left((\lambda t)^{-1}\right), \tag{3.107}$$

or in the limit:

$$\lim_{\lambda t \to \infty} \phi_{Y_\lambda(t)} = \exp\left(-\frac{t^2}{2}\right), \tag{3.108}$$

which is the characteristic function of the standardized Gaussian, implying that $N(t)$ converges in distribution to a Gaussian variable with mean and variance given by λt. In practice, when λt is sufficiently large (say $\lambda t > 1000$), the Gaussian distribution is an excellent approximation for the Poisson distribution. A visual inspection of Figure 3.6 suggests (for t normalized at 1) that when $\lambda = 10$, the distribution is already relatively close to a Gaussian distribution.

After considering the case of a Delta distribution for the GW signal, let us now assume that individual GW signals are independent and identically distributed Uniform distributions taking on values in the normalized $[0, 1]$ interval, and we denote by S the generic random variable that has the same distribution as each one of the S_i variables. We consider again the random sum representing the aggregate signal of astrophysical

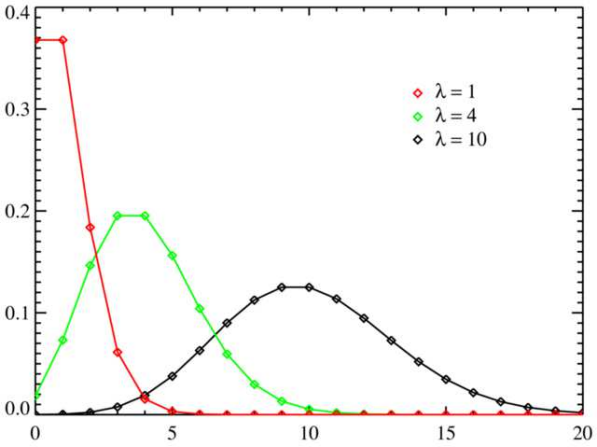

Figure 3.6: Poisson distribution as a function of the intensity parameter λ for $t = 1$. Note that $\Pr(N = 0) = e^{-\lambda}$ and $\Pr(N = 1) = e^{-\lambda}\lambda$.

origin in the period $[0, t]$:

$$S_N(t) = \sum_{i=1}^{N(t)} S_i. \tag{3.109}$$

When N is fixed, the sum of N independent and identically distributed random variables with (continuous) uniform distribution on [0,1] has a distribution called the Irwin-Hall distribution [99], whose density function is given by:

$$f_{IH}(x) = \frac{1}{2(n-1)!} \sum_{k=0}^{N} (-1)^k \begin{pmatrix} N \\ k \end{pmatrix} (x-k)^{N-1} \, sgn(x-k). \tag{3.110}$$

For the uniform distribution, we have $\mu_1 = \frac{1}{2}$ and $\mu_2 = \frac{1}{3}$. It can be shown that the compound Poisson distribution converges to a Gaussian distribution with mean $\frac{\lambda t}{2}$ and variance $\frac{\lambda t}{3}$ as the intensity λ increases for a fixed t, or as t increases for a fixed λ. The speed of convergence can be estimated numerically. To do so, we calculate and plot in Figures 3.7 and 3.8, respectively, the skewness and kurtosis as a function of the intensity parameter λ (taken over a large but limited range extending from 0 to 100) when each signal S follows respectively a delta distribution taking on a value normalized at 1, and the case when each signal follows a uniform distribution taking on value in the interval $[0, 1]$. These quantities can be computed numerically, or calculated using

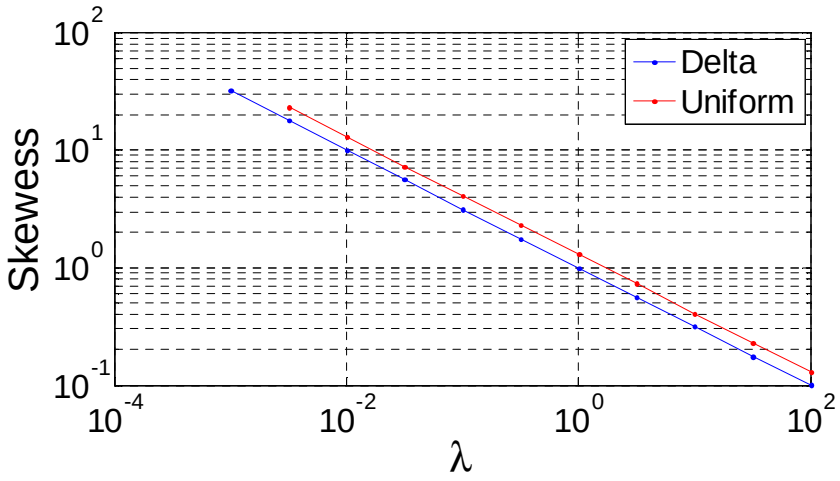

Figure 3.7: Skewness of the compound Poisson process as a function of the intensity λ when the underlying distribution is a Delta or a Uniform distribution.

the analytical expressions below for the skewness Sk and kurtosis Kr of the compound Poisson distribution are [134]:

$$Sk \equiv \frac{\mathbb{E}\left[\left(\sum_{i=1}^{N(t)} S_i\right)^3\right]}{\left(\mathbb{E}\left[\left(\sum_{i=1}^{N(t)} S_i\right)^2\right]\right)^{\frac{3}{2}}} = \frac{8\lambda^3 t^3 \mu_1^3 + 6\lambda^2 t^2 \mu_1 \mu_2 + \lambda t \mu_3}{\left(4\lambda^2 t^2 \mu_1^2 + \lambda t \mu_2\right)^{\frac{3}{2}}} \tag{3.111}$$

$$Kr \equiv \frac{\mathbb{E}\left[\left(\sum_{i=1}^{N(t)} S_i\right)^4\right]}{\left(\mathbb{E}\left[\left(\sum_{i=1}^{N(t)} S_i\right)^2\right]\right)^2} - 3 = \frac{8\lambda^2 t^2 \mu_1 \mu_3 - 32\lambda^4 t^4 \mu_1^4 + \lambda t \mu_4}{\left(4\lambda^2 t^2 \mu_1^2 + \lambda t \mu_2\right)^2}, \tag{3.112}$$

where μ_1, μ_2, μ_3 and μ_4 are, respectively, the first, second, third and fourth moment of the signal distribution S.

From Figures 3.7 and 3.8, we can analyze the convergence to the Gaussian case as the number of underlying sources increases. In the case of a delta distribution, the skewness and kurtosis values are extremely high when the average number of sources λ is lower than 1. For example, when the average number of sources λ is equal to 0.1, the skewness

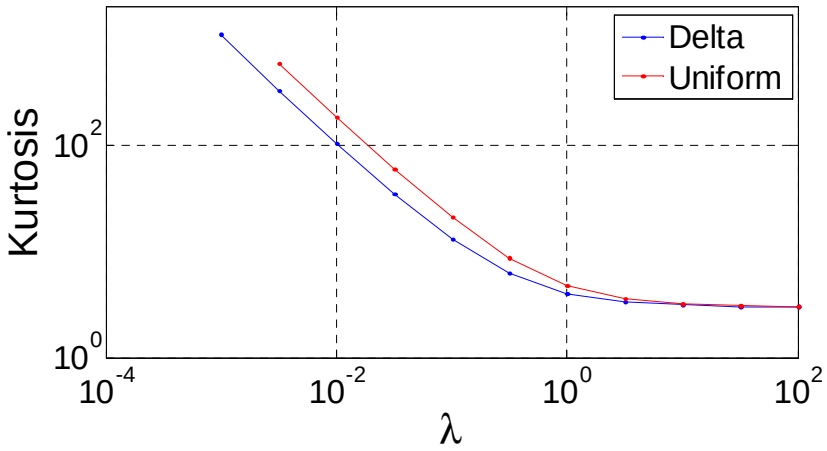

Figure 3.8: Kurtosis of the compound Poisson process as a function of the intensity λ when the underlying distribution is a Delta or a Uniform distribution.

and kurtosis are 3.163 and 13.007 respectively. When the average number of sources λ is equal to 1, the skewness and kurtosis are 1 and 4.003 respectively, while they are 0.316 and 3.099 when $\lambda = 10$ and 0.1 and 3.01 when $\lambda = 100$. The case of a uniform distribution shows a similar pattern, with slightly higher values. For example, when the average number of sources λ is equal to 0.1, the skewness and kurtosis are 4.11 and 21.012 respectively. When the average number of sources λ is equal to 1, the skewness and kurtosis are 1.298 and 4.796 respectively, while they are 0.41 and 3.179 when $\lambda = 10$ and 0.131 and 3.018 when $\lambda = 100$. From this analysis we may arbitrarily set at $\lambda = 10$ the number of overlapping sources required to generate a SGWB signal that can be regarded as "sufficiently" non-Gaussian. From Figure 3.5, we find that with the masses consistent with the recent detections, there will be on average less than 1 binary black hole system emitting in the Advanced LIGO 10-200 Hz band. The number of binary neutron star systems is substantially higher, around 10 in the lowest range of the frequency band. On the other hand, for $f = 10^{-4}$ the number of overlapping sources is extremely large and well within the Gaussian range. We should also keep in mind that Figures 3.7 and 3.8 are based on the assumption of a finite number of i.i.d. sources, while actual sources will have different probability distributions, thus a priori suggesting a higher degree of

non-normality.

Let us finally note that various formal statistical tests can be applied to assess the degree of non-normality of a given SGWB distribution, including the Cramér-von Mises criterion [52], the Kolmogorov-Smirnov test [161], the Shapiro–Wilk test [161], the Anderson-Darling test [20], the Pearson's chi-squared test [42], the Lilliefors test [116], the D'Agostino's K-squared test [54], and the Jarque-Bera test [104]. The Jarque-Bera (JB) test is particularly well-suited for our analysis, since it is explicitly derived from sample *skewness* and *kurtosis* estimates, which will be the focus of our attention in chapters 4 and 5. The JB test statistic is defined as [104]:

$$JB = \frac{n}{6}\left(Sk^2 + \frac{1}{4}\widetilde{Kr}^2\right), \tag{3.113}$$

where n is the number of observations, Sk is the skewness and $\widetilde{Kr}$ the excess kurtosis of the signal distribution. The statistic JB has an asymptotic χ^2 (chi-square) distribution with two degrees of freedom (which is the distribution followed by the sum of two i.i.d. squared standardized Gaussian random variables) and can be used to test the null hypothesis that the data are from a normal distribution. The null hypothesis is a joint hypothesis of the skewness being zero and the excess kurtosis being 0, so that any deviation from this increases the JB statistic. When applying the formal JB test to the compound Poisson process with an underlying distribution that is either delta or uniform, we find that the Gaussian assumption is strongly rejected given the skewness and kurtosis parameters displayed in Figures 3.7 and 3.8, and this for all tested values of the intensity process λ in the range $[10^{-4}, 100]$.

3.3.2 Non-Gaussian SGWB Distributions and the Edgeworth Expansion

As discussed above, there are strong reasons to believe that the GW background from astrophysical sources will fall in the non-Gaussian regime, which raises the question of

a suitable extension of the detection methodologies to account for the presence of these deviations from the Gaussian assumption. Broadly speaking, there are 3 main approaches to statistical problems involving non-Gaussian distributions. The first approach is the parametric approach, which consists of assuming a given non-Gaussian signal distribution (Student, inverse Gaussian, Pareto-Levy, etc.), and using the assumed density to derive, subject to analytical tractability, an expression for the log likelihood, which can subsequently be used to perform parameter estimation. The main limitation with the parametric approach is specification risk, namely the fact that the assumed non-Gaussian distribution may not be a good fit for the unknown signal distribution. Besides, it is in general impossible to obtain the likelihood function (frequentist approach) or the marginalized likelihood function (Bayesian approach) in analytical form for non-Gaussian densities. For example, a Student's t-distribution model is considered in [49] as an example of a non-Gaussian parameterized SGWB signal prior, but the authors do not use this model in their analysis because of a lack of tractability in their Bayesian approach,as can be seen from the following quote: "Also, as we shall discuss further in Sec. V, we do not consider a signal+noise model with a non-standardized Student's t-distribution for the non-Gaussian stochastic gravitational-wave component. This is because of the computational costs associated with the marginalized likelihood evaluations (see Eq. (18)), which are needed for the Bayesian model selection calculations." The authors also make the following statement when analyzing the marginalized likelihood function: "Unfortunately, we do not know how to analytically evaluate such an integral. (...) Thus, if we want to use this distribution as one of our non-Gaussian signal models, we would need to evaluate the above integrals numerically."

To alleviate the concern over specification risk (i.e., the risk that the true unknown distribution differs from the assumed distribution), one may instead prefer a non parametric approach, where no assumption is made about the unknown distribution, and where sample information is used to perform parameter estimation. Two main shortcomings of this approach are the fact that it does not allow for likelihood maximization, and

also the lack of robustness due to the sole reliance on sample-based information. Finally, one interesting alternative to these two extreme approaches exists, which is known as a semi-parametric approach, where the focus is on approximating the unknown signal distribution as a transformation of a reference function (typically the Gaussian density). This can be done formally through the Edgeworth expansion, which is the most straightforward approximation of a non-Gaussian distribution, and which provide an explicit expression of the rate of convergence of the central limit theorem. We subsequently consider how it can be applied to the detection of non-Gaussian SGWBs in the course of next chapter. We will actually show in chapter 4 then that the use of the Edgeworth expansion not only addresses the concern over specification risk, but also the concern over lack of tractability. This is because the Edgeworth expansion involves an approximate density than can be written as the product of a polynomial and the Gaussian density, and standard results on Gaussian integrals can be used to calculated the log-likelihood in closed-form (see for example equation 4.51).[9]

More precisely, the Edgeworth expansion represents an approximation to the density of a normalized sum of i.i.d. copies of X the distribution of which satisfies some regularity conditions involving existence of moments and an integrability condition of the characteristic function (chf) of X. The integrability of the chf guarantees uniform convergence of the approximating density to the limit density while the existence of higher-order moments guarantees a faster rate of convergence to the limit density. Here we provide a result regarding the convergence rate of the Edgeworth expansion. This is a standard result, for which we provide a proof adapted from [75]. The reminder of this section can be regarded as a pure mathematical development that contains no original result. This material merely serves the purpose of providing the reader with a formal proof of

[9] This approach, which we introduce for the first time in the analysis of SGWB signals in [121], was also considered in [49]: "It is possible to consider an Edgeworth expansion of the Student t-distribution in terms of its non-zero cumulants, c_2, c_4, etc. But then truncating the expansion after a nite number of terms would produce a different non-Gaussian distribution, that would behave differently in model comparison tests from the full Student's t-distribution." Since there is no a priori justification for the use of the Student's t-distribution, we argue that an approximative version of that distribution classically involving cumulants up to the order 4 would be a natural tractable alternative.

the Edgeworth expansion, which is heavily used in chapters 4 and 5 for the analysis of SGWB signals.

We first introduce some notation. Let $X_1, \ldots, X_n$ denote i.i.d. copies of X and denote by $f_n(x)$ the density of $S_n = (X_1 + \ldots + X_n)/\sqrt{n}$ in which $\mathbb{E}(X^2) = 1$. Without loss of generality, we have also assumed $\mathbb{E}(X) = 0$. Further on, denote by $\mu_r = \mathbb{E}(X^r)$ the r-th moment of X, by $\varphi(t) = \mathbb{E}\left(e^{itX}\right)$ the chf of X, and by $\phi(x) = e^{-x^2/2}/\sqrt{2\pi}$ the density of the centered standard normal distribution.

Proposition 2 *Suppose that the moments $\mu_3, \ldots, \mu_r$ exist and that $|\varphi|^\nu$ is integrable for some $\nu \geq 1$. Then, $f_n(x)$ exists for $n \geq \nu$ and as $n \to \infty$:*

$$f_n(x) = \phi(x) + \phi(x) \sum_{k=3}^{r} \frac{P_{k-2}(x)}{n^{k/2-1}} + O(n^{-r/2+1}). \tag{3.114}$$

Here $P_k(x)$ is a real polynomial depending only on $\mu_3, \ldots, \mu_k$ but not on n or r.

Proof. We provide a sketch of the proof in which the particular form of the polynomials is derived. Denote by $c_0 = c_1 = 0$, $c_2 = 1$, $c_3, \ldots, c_r$ the cumulants of the X. The logarithm of the chf of X can be expanded:

$$\begin{aligned} \psi(t) = \log \varphi(t) &= \sum_{j=0}^{r} \frac{c_j (it)^j}{j!} + O(|t|^r) \\ &= -\frac{t^2}{2!} + \frac{c_3 (it)^3}{3!} + \ldots + \frac{c_r (it)^r}{r!} + O(|t|^r) \end{aligned} \tag{3.115}$$

in a neighborhood of $t = 0$. The chf of the normalised sum can be represented through the chf of the summands which are independent copies of X:

$$\varphi_n(t) = E \exp\left(it \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \right) = \varphi^n\left(\frac{t}{\sqrt{n}}\right).$$

It is more convenient to work with the logarithm where we can plug in the Taylor expan-

sion in (3.115):

$$\begin{aligned}\psi_n(t) = \log \varphi_n(t) &= n\psi\left(\frac{t}{\sqrt{n}}\right) \\ &= -\frac{t^2}{2} + \frac{c_3(it)^3}{3!n^{3/2-1}} + \ldots + \frac{c_r(it)^r}{r!n^{r/2-1}} + O(|t|^r) \\ &= -\frac{t^2}{2} + \psi_{n,r}(it) + O(|t|^r).\end{aligned}$$

where

$$\psi_{n,r}(x) = \frac{c_3 x^3}{3!n^{3/2-1}} + \ldots + \frac{c_r x^r}{r!n^{r/2-1}}. \tag{3.116}$$

As a result, for the chf of the normalised sum S_n we get:

$$\begin{aligned}\varphi_n(t) = \exp \psi_n(t) &= \exp\left(-\frac{t^2}{2}\right) \exp(\psi_{n,r}(t) + O(|t|^r)) \\ &= \exp\left(-\frac{t^2}{2}\right) \exp(\psi_{n,r}(it)) + O(|t|^r).\end{aligned} \tag{3.117}$$

At this stage, the chf of the normalised sum has a Gaussian component multiplied by a term which is an exponent of a polynomial of it. The goal now is to write an expansion for the factor. Since we have assumed existence of moments up to the order of r, then the chf of the sum can be expanded in a manner similar to (3.115) and the Taylor expansion would be of the r-th order. In (3.117), however, the Gaussian component is of the second order, i.e. $\exp(-t^2) \approx 1 - t^2$ in a neighborhood of $t = 0$, and therefore the factor will be of order $r - 2$. We use the power series expansion of the exponent to arrive at:

$$\varphi_n(t) = \exp\left(-\frac{t^2}{2}\right)\left(\sum_{j=0}^{r-2} \frac{(\psi_{n,r}(it))^j}{j!}\right) + O(|t|^r). \tag{3.118}$$

The next step is to raise the polynomials to the corresponding powers in the expansion and to order the terms according to the power of n. Note that the polynomials $\psi_{n,r}(x)$ defined in (3.116) are of order r. Raising them to a power in (3.118) will produce terms of higher order which we can safely ignore because the order of the approximation is r.

Thus, the goal now is to write down explicitly the terms $Q_j(x)$ in the representation:

$$\varphi_n(t) = \exp\left(-\frac{t^2}{2}\right)\sum_{j=0}^{r-2}\frac{Q_j(it)}{n^{j/2}} + O(|t|^r). \tag{3.119}$$

which, as explained above, are polynomials. We proceed on a term by term basis. $Q_0(x) = 1$ which is obvious. $Q_1(x)$ collects the terms that contain $n^{1/2}$ in the denominator. There is only one such term which is the first term of $\psi_{n,r}(x)$ where the order is as given in (3.116), i.e. $Q_1(x) = \frac{c_3x^3}{3!}$. Next, $Q_2(x)$ collects all terms that have n in the denominator. There are two such terms – the second one in $\psi_{n,r}(x)$ and the first one in $(\psi_{n,r}(x))^2$, i.e. $Q_2(x) = \frac{c_4x^4}{4!} + \frac{(c_3)^2x^6}{2!(3!)^2}$. We can proceed in this way with the higher order terms. The first few expressions are summarised below:

$$\begin{aligned}
Q_0(x) &= 1\\
Q_1(x) &= \frac{c_3x^3}{3!}\\
Q_2(x) &= \frac{c_4x^4}{4!} + \frac{(c_3)^2x^6}{2!(3!)^2}\\
Q_3(x) &= \frac{c_5x^5}{5!} + \frac{c_3c_4x^7}{3!4!} + \frac{(c_3)^3x^9}{(3!)^3}.
\end{aligned} \tag{3.120}$$

So far, we have managed to derive an approximation of the chf of the normalised sum. Denote the approximation by:

$$\widetilde{\varphi}_n(t) = \exp\left(-\frac{t^2}{2}\right)\sum_{j=0}^{r-2}\frac{Q_j(it)}{n^{j/2}}. \tag{3.121}$$

The next step is to translate it into a density approximation of the normalised sum. To this end, we employ the inversion formula:

$$\widetilde{f}_n(t) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\widetilde{\varphi}_n(t)dt.$$

The rationale for this is the inequality:

$$|f_n(x) - \widetilde{f}_n(x)| \leq \frac{1}{2\pi}\int_{\mathbb{R}} |e^{-itx}||\varphi_n(t) - \widetilde{\varphi}_n(t)|dt = \frac{1}{2\pi}\int_{\mathbb{R}} |\varphi_n(t) - \widetilde{\varphi}_n(t)|dt.$$

The right hand-side converges to zero because of the assumed integrability of the chf for $n \geq \nu$. Further on, since it does not depend on x, the convergence in terms of densities is uniform. The rate of convergence is derived from the Fourier norm but we are not going to do it here.

From the inversion formula, we get:

$$\begin{aligned}\widetilde{f}_n(t) &= \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\left(e^{-\frac{t^2}{2}}\sum_{j=0}^{r-2}\frac{Q_j(it)}{n^{j/2}}\right)dt \\ &= \sum_{j=0}^{r-2}\frac{1}{n^{j/2}2\pi}\int_{\mathbb{R}} e^{-itx}\left(e^{-\frac{t^2}{2}}Q_j(it)\right)dt.\end{aligned} \tag{3.122}$$

Since $Q_j(x)$ is a polynomial, the generic form of the terms involving the integral is

$$\frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\left(e^{-\frac{t^2}{2}}(it)^j\right)dt = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\frac{t^2}{2}}\left(e^{-itx}(it)^j\right)dt.$$

The term in the parentheses can be represented through the derivatives of e^{-itx} with respect to x. Indeed, it can be easily verified that $(-1)^j\frac{d^j}{dx^j}e^{-itx} = e^{-itx}(it)^j$. As a consequence, for the generic term we get:

$$\begin{aligned}\frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\left(e^{-\frac{t^2}{2}}(it)^j\right)dt &= \frac{(-1)^j}{2\pi}\int_{\mathbb{R}} e^{-\frac{t^2}{2}}\left(\frac{d^j}{dx^j}e^{-itx}\right)dt \\ &= \frac{(-1)^j}{2\pi}\frac{d^j}{dx^j}\int_{\mathbb{R}} e^{-\frac{t^2}{2}}e^{-itx}dt \\ &= (-1)^j\frac{d^j}{dx^j}\phi(x).\end{aligned}$$

where the last equality follows by recognizing the Gaussian chf in the $e^{-\frac{t^2}{2}}$ factor and

applying the inversion formula to it. The expression can be further simplified:

$$\frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\left(e^{-\frac{t^2}{2}}(it)^j\right)dt = \phi(x)H_j(x), \tag{3.123}$$

where $H_j(x)$ denotes the polynomial remaining after differentiating the Gaussian density. These polynomials are known as Hermite polynomials.

The Hermite polynomials are a classical orthogonal polynomial sequence, which are used in a variety of context in mathematics and physics, which have been introduced by Laplace in 1820.

We are ready to derive an expression for the density. Denote by $Q_j(x) = \sum_k b_{jk}x^k$ and take advantage of the derived expression for the generic term in (3.123). Then, the expression in (3.122) becomes:

$$\begin{aligned}\widetilde{f}_n(t) &= \sum_{j=0}^{r-2}\frac{1}{n^{j/2}}\sum_k b_{jk}\frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\left(e^{-\frac{t^2}{2}}(it)^k\right)dt \\ &= \sum_{j=0}^{r-2}\frac{1}{n^{j/2}}\phi(x)\sum_k b_{jk}H_k(x)dt \\ &= \phi(x)\sum_{j=0}^{r-2}\frac{P_j(x)}{n^{j/2}}dt.\end{aligned} \tag{3.124}$$

in which the polynomial $P_j(x)$ is obtained from $Q_j(x)$ by simply replacing the term x^k with $H_k(x)$. Noticing that $P_0(x) = 1$ and that the index k in the theorem is related to j according to $k = j + 2$ completes the proof. ■

Using the derived expressions for the polynomial $Q_j(x)$ in (3.120), the first few terms of $P_k(x)$ expressed in terms of the cumulants are given by [75]:

$$P_3(x) = \frac{c_3H_3(x)}{3!} \tag{3.125}$$

$$P_4(x) = \frac{c_4H_4(x)}{4!} + \frac{(c_3)^2H_6(x)}{2!(3!)^2} \tag{3.126}$$

$$P_5(x) = \frac{c_5H_5(x)}{5!} + \frac{c_3c_4H_7(x)}{3!4!} + \frac{(c_3)^3H_9(x)}{(3!)^3} \tag{3.127}$$

where $H_k(x)$ stands for the Hermite polynomial defined by:

$$\frac{d^k}{dx^k}\phi(x) = (-1)^k H_k(x)\phi(x) \tag{3.128}$$

The first few Hermite polynomials used in the expressions above are given by:

$$H_3(x) = x^3 - 3 \tag{3.129}$$

$$H_4(x) = x^4 - 6x^2 + 3 \tag{3.130}$$

$$H_5(x) = x^5 - 10x^3 + 15x \tag{3.131}$$

$$H_6(x) = x^6 - 15x^4 + 45x^2 - 15 \tag{3.132}$$

$$H_7(x) = x^7 - 21x^5 + 105x^3 - 105x \tag{3.133}$$

$$H_9(x) = x^9 - 36x^7 + 378x^5 - 1260x^3 + 945x \tag{3.134}$$

Using the link between cumulants and moments, the expressions for P_3, $P_4$and P_5 can be re-stated in terms of the moments of X. It should also be noted that if we assume that X has a bounded continuous density, then $\nu = 1$. The converse statement also holds. In other words, the existence of continuous density of X guarantees integrability of the chf of X. The derivation relies on the assumed zero mean and unit variance of X. The case of non-nonrmalized X can be handled directly through the normalized case. First, suppose that $\mathbb{E}(X) = \mu$. Then, the sum $S_n = (X_1 + \ldots + X_n)/\sqrt{n}$ diverges because it is not properly normalized. Indeed, with some abuse of notation:

$$S_n = \frac{(X_1 + \ldots + X_n)}{n}\sqrt{n} \to \mu\sqrt{n} \to \infty, \tag{3.135}$$

where the limit is in almost sure sense (meaning that it holds with probability 1) and the first part is due to the strong law of large numbers. The convergence in distribution would follow if the sum is centered by subtracting the mean of the sum. Thus, as far

as the Edgeworth approximation goes, we can either assume that $\mathbb{E}(X) = 0$ and derive the approximation of the density of the sum $S_n = (X_1 + \ldots + X_n)/\sqrt{n}$ or if we assume $\mathbb{E}(X) = \mu$, then we should consider $S_n^* = (X_1 + \ldots + X_n - n\mu)/\sqrt{n}$ which means that, again, the sum has a zero mean after centering. Second, assume that $\mathbb{E}(Y) = 0$ and $\mathbb{E}(Y^2) = \sigma^2$. The sum $S_n = (Y_1 + \ldots + Y_n)/\sqrt{n}$ has a zero mean and a variance $\mathbb{E}(S_n^2) = \sigma^2$. Set $Y = \sigma X$, where $\mathbb{E}(X^2) = 1$. The logarithms of the chfs of Y and X are related through a simple formula:

$$\psi_Y(t) = \psi_X(\sigma t), \tag{3.136}$$

which implies that the cumulants of X and Y are related through $c_{X,j} = c_{Y,j}/\sigma^j$, $j \geq 1$. The density of $S_n = (Y_1 + \ldots + Y_n)/\sqrt{n}$ can be approximated using (3.114). The following corollary holds.

Corollary 3 *Suppose that the random variable Y is such that $\mathbb{E}(Y) = 0$, $\mathbb{E}(Y^2) = \sigma^2$ and $\mathbb{E}(|Y|^r) < \infty$ for $r \geq 2$. Denote the cumulants of Y by $c_{Y,j}$ for $3 \leq j \leq r$. The Edgeworth approximation of the density of $S_n = (Y_1 + \ldots + Y_n)/\sqrt{n}$ equals:*

$$\widetilde{f}_n(x) = \frac{1}{\sigma}\phi\left(\frac{x}{\sigma}\right)\left(1 + \sum_{k=3}^{r}\frac{P_k(x/\sigma)}{n^{k/2-1}}\right), \tag{3.137}$$

where the cumulants corresponding to $P_k(x)$ are equal to $c_j = c_{Y,j}/\sigma^j$ for $3 \leq j \leq r$.

Proof. There are two equivalent ways to arrive at the result in (3.137). The analytic approach would be to repeat the arguments in the proof of Theorem 2 noting that

$$\begin{aligned}\widetilde{\varphi}_n(t) &= \exp\left(-\frac{\sigma^2 t^2}{2}\right)\exp\left(\frac{c_{Y,3}(it)^3}{3!n^{3/2-1}} + \ldots + \frac{c_{Y,r}(it)^r}{r!n^{r/2-1}}\right), \\ &= \exp\left(-\frac{\sigma^2 t^2}{2}\right)\sum_{j=0}^{r-2}\frac{Q_j(it)}{n^{j/2}}\end{aligned}$$

where $Q_j(x)$ is as in (3.120) with $c_j = c_{Y,j}$. Further on, it can be checked directly that

$$\frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\left(e^{-\frac{t^2}{2}}(it)^j\right)dt = \frac{(-1)^j}{\sigma}\frac{d^j}{dx^j}\phi(x/\sigma) = \frac{1}{\sigma}\frac{1}{\sigma^j}\phi(x/\sigma)H_j(x/\sigma)$$

and plugging everything into (3.124) we get

$$\widetilde{f}_n(t) = \frac{1}{\sigma}\phi(x/\sigma)\sum_{j=0}^{r-2}\frac{P_j^*(x/\sigma)}{n^{j/2}}dt.$$

where the polynomial $P_j^*(x)$ is obtained from $Q_j(x)$ by replacing the term x^k with $H_k(x)/\sigma^k$. Note that the term σ^k can be absorbed by the cumulants of Y. Thus, $P_j^*(x)$ turns out to be the same as $P_j(x)$ in which $c_j = c_{Y,j}/\sigma^j$.

A probabilistic proof would be more straightforward. Denote by $S_n^* = (Y_1 + \ldots + Y_n)/\sigma\sqrt{n} = S_n/\sigma$. The sum S_n^* is now normalised and according to Theorem 2 its Edgeworth approximation reads

$$\widetilde{f}_n^*(x) = \phi(x)\left(1 + \sum_{k=3}^{r}\frac{P_k(x)}{n^{k/2-1}}\right)$$

where the cumulants in the polynomials $P_k(x)$ are those of $X = Y/\sigma$. We can now write down the density of the random variable $S_n = \sigma S_n^*$

$$\widetilde{f}_n(x) = \widetilde{f}_n^*(x/\sigma)/\sigma$$

and we arrive at (3.137). ■

We now present a number of concluding comments about the Edgeworth expansion. Let us first denote the right-hand side of (3.114) it by $\widetilde{f}_{n,r}(x)$:

$$\widetilde{f}_{n,r}(x) = \phi(x) + \phi(x)\sum_{k=3}^{r}\frac{P_k(x)}{n^{k/2-1}}. \tag{3.138}$$

The expression implies that a correction factor is applied to the Gaussian density. The

function $\widetilde{f}_{n,r}(x)$, however, is not necessarily a density function itself; it can take negative values, a problem which we revisit in the next chapter, where we argue that the problem does not arise for the examples that we analyze (see Figures 4.3, 4.4, 4.5 and 4.6), and where we also suggest to take the squared value of the correction term should the problem arise in practice (equation 4.44). On the other hand, the function $\widetilde{f}_{n,r}(x)$ does integrate to 1. To see this, note that the approximation of the chf in the derivation is constructed such that $\widetilde{\varphi}_n(0) = 1$ for any n. From the transform $\widetilde{\varphi}_n(t) = \int_{\mathbb{R}} e^{itx}\widetilde{f}_n(x)dx$, it follows that:

$$1 = \widetilde{\varphi}_n(0) = \int_{\mathbb{R}} \widetilde{f}_n(x)dx. \tag{3.139}$$

Using Corollary 3, the expansion for the special case of a symmetric X such that $\mathbb{E}(X^2) = \sigma^2$ and assuming finite moments up to the fourth order is given by:

$$\widetilde{f}_{n,4}(x) = \phi(x/\sigma)/\sigma + \phi(x/\sigma)\frac{\mu_4/\sigma^4 - 3}{24n\sigma}((x/\sigma)^4 - 6(x/\sigma)^2 + 3). \tag{3.140}$$

The expression in the numerator, $\mu_4/\sigma^4 - 3$, is the kurtosis of X. The convergence rate in (3.114) can be made more precise.

Proposition 4 *Let $X_1, X_2, \ldots, X_n$ be i.i.d. copies of X with $\mathbb{E}(X) = 0$ and $\mathbb{E}(X^2) = 1$ and $\mathbb{E}(|X|^r) < \infty$ for some integer $r \geq 2$. Suppose that $|\varphi|^\nu$ is integrable for some ν. Let $\widetilde{f}_{n,r}(x)$ denote the Edgeworth approximation in* (3.114)*, then:*

$$\sup_x (1 + |x|^r)|f_n(x) - \widetilde{f}_{n,r}(x)| = O(n^{-r/2+1}). \tag{3.141}$$

Let us emphasize that the idea behind Edgeworth expansion can be applied directly to a given random variable X, not necessarily in the context of sums of i.i.d. copies of X, which then leads to the so-called Gram-Charlier approximation. In the setting of CLT, the value of the Edgeworth expansion is that it provides a convergence rate with respect to n. Outside of the context of the CLT, it is just an approximation problem – it approximates the density of X through a Gaussian density and a correction term. This

will be the case in our analysis of a non-Gaussian noise distribution in chapter 5. In fact, the arguments behind the proof of Theorem 2 can be repeated without modification and this is what we do next. Suppose that X has a zero mean and unit variance and that $\mathbb{E}\left(|X|^r\right) < \infty$ for some $r > 2$. Then, for the chf of X we can write down:

$$\begin{aligned}
\varphi_X(t) &= \exp\left(-\frac{t^2}{2}\right)\exp\left(\frac{c_{X,3}(it)^3}{3!}+\ldots+\frac{c_{X,r}(it)^r}{r!}\right)+O(|t|^r) && (3.142)\\
&= \exp\left(-\frac{t^2}{2}\right)\exp\left(\psi_r(it)\right)+O(|t|^r) && (3.143)\\
&= \exp\left(-\frac{t^2}{2}\right)\left(\sum_{j=0}^{r}\frac{(\psi_r(it))^j}{j!}\right)+O(|t|^r) && (3.144)\\
&= \exp\left(-\frac{t^2}{2}\right)\sum_{j=0}^{r-2}\alpha_{j,r}(it)^j+O(|t|^r) && (3.145)
\end{aligned}$$

where $\alpha_{j,r}$ collect the terms multiplying the corresponding power of (it). Applying the Fourier inversion formula, we end up with the Gram-Chalier decomposition (see next chapter for more details):

$$\begin{aligned}
\widetilde{f}_X(t) &= \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\widetilde{\varphi}_X(t)dt && (3.146)\\
&= \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}e^{-t^2/2}\sum_{j=0}^{r-2}\alpha_{j,r}(it)^j dt && (3.147)\\
&= \sum_{j=0}^{r-2}\alpha_{j,r}\frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}e^{-t^2/2}(it)^j dt && (3.148)\\
&= \phi(x)\sum_{j=0}^{r-2}\alpha_{j,r}H_j(x) && (3.149)
\end{aligned}$$

The main focus of the next chapter is to use the 4th-order Edgeworth expansion of the unknown density for the signal and noise distributions to obtain an explicit expression for the nearly optimal likelihood statistic in the non-Gaussian case, and explore the implications in terms of detection of non-Gaussian SGWB and estimation of the parameters of the SGWB distribution.

4

A Semi-Parametric Approach to the Detection of Non-Gaussian Stochastic Gravitational Wave Backgrounds

---

This chapter is an expanded version of the following paper [121]: "Semiparametric approach to the detection of non-Gaussian gravitational wave stochastic backgrounds", L. Martellini and T. Regimbau, 2014, Physical Review D, 89, 12, 124009.

As recalled in the previous chapter, the optimal detection strategy to search for a stochastic background is to cross correlate the output of two detectors (or of a network of detectors) to eliminate the instrumental noise [15]. The GW background is usually assumed to be Gaussian invoking the central limit theorem, and thus completely characterized by its mean and variance. However recent predictions based on population modeling suggest that for many astrophysical models, there may not be enough overlapping sources, resulting in the formation of a non-Gaussian background. As also recalled in the previous chapter, it has also been shown that the background from cosmic strings could be dominated by a non-Gaussian contribution arising from some of the closest sources [55, 144] that would generate GW signals that would not be strong enough to be

individually detectable. The identification of a non-Gaussian signature would not only permit one to distinguish between astrophysical and cosmological GW backgrounds and gain confidence in a detection, but the measurement of extra information should also help impose additional constraints on parameters of underlying astrophysical and cosmological models.

In the past decade a few methods have been proposed to search for a non-Gaussian stochastic background, including the *probability horizon* concept developed by [50] based on the temporal evolution of the loudest detected event on a single detector, the maximum likelihood statistic of [62] which extends the standard cross correlation statistic in the time domain in the case of short astrophysical bursts separated by long periods of silence, the fourth-order correlation method [157] which uses fourth-order correlation between four detectors to measure the third and the fourth moments of the distribution of the GW signal, or the recent extension of the standard cross-correlation statistic by [168].

In this chapter, we start from the general formalism presented in [62] to analyze small deviations from the Gaussian distribution. In this case the cross-correlation statistic is almost optimal and allows for the estimation of the variance of the signal distribution, but it cannot be used to estimate higher order moments. The approach we propose is based on the Edgeworth expansion, which is a formal asymptotic expansion of the characteristic function of the signal distribution, whose unknown probability density function is to be approximated in terms of the characteristic function of the Gaussian distribution.

The Edgeworth expansion has been used in many applications in engineering, economics and finance, and it has also been successfully used in astrophysics, in particular in the analysis of the microwave background anisotropies ([154], [111], [108], [129], [77], [83], [19]) and in the analysis of the velocity distributions and structure of elliptical galaxies ([148], [171], [84], [93], [30]). To the best of our knowledge, however, the paper [121] from which this chapter is drawn is the first attempt to use the Edgeworth expansion in the context of analysis and detection of stochastic gravitational wave backgrounds. Since the Edgeworth expansion provides asymptotic correction terms to the Central Limit The-

orem up to an order that depends on the number of moments available, it is ideally suited for the analysis of stochastic gravitational wave backgrounds generated by a finite number of astrophysical sources. It is also well-suited for the analysis of signals from cosmological origin in case the deviations from the Gaussian assumption are not too strong. Using a 4th-order Edgeworth expansion, we obtain an explicit expression for the nearly optimal non-Gaussian likelihood statistic, which is shown to involve standard Gaussian integrals. This expression generalizes the standard maximum likelihood detection statistic, which is recovered in the limit of vanishing third and fourth cumulants of the empirical conditional distribution of the detector measurement. We use numerical procedures to generate maximum likelihood estimates for the gravitational wave distribution parameters for a set of heavy-tailed distributions and find that the fourth cumulant can be estimated with reasonable precision when the ratio between the signal and the noise variances is larger than 0.01. The use of the non-Gaussian detection statistic comes with no loss of sensitivity when the signal variance is small compared to the noise variance, and involves an efficiency gain when the noise and the signal are of comparable magnitudes. The rest of the chapter is organized as follows. In section 4.1, we introduce a detection statistic for a non-Gaussian stochastic background distribution. In section 4.2, we present a number of numerical simulations. In section 4.3, we discuss how the approach can be extended to analyze signals in the popcorn regime.

4.1 Detection Methods for Non-Gaussian Gravitational Wave Backgrounds

Following [62], we consider two gravitational wave detectors, assumed to be identical, which implies that the noise for both detectors is drawn from the same distribution, as well as coincident and coaligned, which also implies that the signal measured by both detectors is identical. Using the same notation as in the previous chapter, we decompose the measurement output h_{it} for detector $i = 1, 2$ at time t in terms of noise n_{it} (specific

to each detector) versus signal s_t (common to both detectors): $h_{it} = n_{it} + s_t$. Here we assume that the noise in detector 1 and 2 follow uncorrelated Gaussian distributions $\mathcal{N}_1$ and $\mathcal{N}_2$ with zero mean and standard deviations denoted by σ_1 and σ_2, respectively. We analyze the case of non-Gaussian noise distributions in chapter 5. We denote by $c_2 \equiv \alpha^2$ the variance of the signal distribution $\mathcal{S}$, and by c_3 and c_4, respectively, the third- and fourth-order cumulants of the signal distribution (see formal definitions below). When the signal is Gaussian, we have that $c_3 = c_4 = 0$. Note that we use capital letters for the random variables, e.g., $\mathcal{S}$ for the signal, and small letters for their realizations given a given random outcome ω. In other words, we write $\mathcal{S}_t(\omega) = s_t$, where $\mathcal{S}_t$ for $t = 1, ..., T$ are identical copies of the stationary random variable $\mathcal{S}$, and where T is the number of data points.[1]

4.1.1 Gram-Charlier and Edgeworth Expansions

Our ambition is to propose a semi-parametric approach that allows one to approximate the unknown density as a transformation of a reference function (typically the Gaussian density), involving higher-order moments/cumulants of the unknown distribution. This approach, formally introduced in the previous chapter, has been heavily used in statistical problems involving a mild departure from the Gaussian distribution; it is generally more robust than the non-parametric approach, since the sample information is only used to generate estimates for the 3rd and 4th cumulants of the unknown distribution function, and it does not suffer from the specification risk inherent to the parametric approach. In what follows, we show that a semi-parametric expansion of the unknown signal density function can be used to obtain an analytical derivation of the nearly optimal maximum likelihood detection statistic for non-Gaussian gravitational wave stochastic backgrounds.

To see this, let us denote by f be the density function of the unknown distribution of the stochastic background signal S which we want to approximate as a function of the

[1] Here T is not the observation time but the number of data points, which is the observation time multiplied by the sampling rate

Gaussian density function ϕ. We first recall the definition for G, the *moment generating function* of S:

$$G_s(t) = E\left[e^{tS}\right] = \int_{-\infty}^{\infty} e^{tx} f_s(x)\, dx, \tag{4.1}$$

which is related to the characteristic distribution ψ, i.e., the Fourier transform of the function f_s, by $\psi_s(t) = G_s(it)$.

Using the Taylor expansion of the exponential function around 0, $e^x = \sum_{j=0}^{\infty} \frac{x^j}{j!}$, we obtain the following expression for the characteristic function:

$$\psi_s(t) = E\left[e^{itS}\right] = \sum_{j=0}^{\infty} \frac{(it)^j}{j!} E\left(S^j\right) \equiv \sum_{j=0}^{\infty} \frac{(it)^j}{j!} \mu_j, \tag{4.2}$$

where μ_j denotes the j^{th} (non central) moment of the distribution of S. From this, we see that j^{th} (non central) moment of the distribution is given by the j^{th} derivative of the moment-generating function G_s taken at $t = 0$ (hence the name *moment generating function*): $\mu_j = G_s^{(j)}(0) = (-i)^j \psi_s^{(j)}(0)$.

We also recall the definition of the *cumulant generating function* g_s as the logarithm of the characteristic function:

$$g_s(t) = \log G_s(t) = \log \sum_{j=1}^{\infty} \frac{t^j}{j!} \mu_j. \tag{4.3}$$

A Taylor expansion of the cumulant generating function g_s can be written under the following form:

$$g_s(t) = g_s(0) + \sum_{j=1}^{\infty} \frac{t^j}{j!} g_s^{(j)}(0), \tag{4.4}$$

and we define $c_j = g_s^{(j)}(0)$ as the j^{th} *cumulant* of the random variable S.

A moments-to-cumulants relationship can be obtained by expanding the exponential and equating coefficients of t^j in:

$$G_s(t) = \exp\left[g_s(t)\right] \Longleftrightarrow \sum_{j=0}^{\infty} \frac{t^j}{j!} \mu_j = \exp\left[\sum_{j=1}^{\infty} \frac{t^j}{j!} c_j\right]. \tag{4.5}$$

Conversely, a cumulants-to-moments relationship is obtained by expanding the logarithmic and equating coefficients of t^j in $g_s(t) = \log G_s(t)$. Hence we have:

$$c_1 = g_s'(0) = \mu_1 = \mu \tag{4.6}$$

$$c_2 = g_s''(0) = \mu_2 - \mu_1^2 = \alpha^2 \tag{4.7}$$

$$c_3 = g_s^{(3)}(0) = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3 \tag{4.8}$$

$$c_4 = g_s^{(4)}(0) = \mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4 \tag{4.9}$$

We note in particular that the first cumulant is equal to the first moment (the mean), and the second cumulant is equal to the second-centered moment (the variance). Cumulants are often simpler than moments. For example, the first 6 moments of a Gaussian distribution with mean μ and standard deviation α are:

$$\mu_1 = \mu \tag{4.10}$$

$$\mu_2 = \mu^2 + \alpha^2 \tag{4.11}$$

$$\mu_3 = \mu^3 + 3\mu\alpha^2 \tag{4.12}$$

$$\mu_4 = \mu^4 + 6\mu^2\alpha^2 + 3\alpha^4 \tag{4.13}$$

$$\mu_5 = \mu^5 + 10\mu^3\alpha^2 + 15\mu\alpha^4 \tag{4.14}$$

$$\mu_6 = \mu^6 + 15\mu^4\alpha^2 + 45\mu^2\alpha^4 + 15\alpha^6 \tag{4.15}$$

$$\mu_7 = \mu^7 + 21\mu^5\alpha^2 + 105\mu^3\alpha^4 + 105\mu\alpha^6 \tag{4.16}$$

$$\mu_8 = \mu^8 + 28\mu^6\alpha^2 + 210\mu^4\alpha^4 + 420\mu^2\alpha^6 + 105\alpha^8 \tag{4.17}$$

while the cumulants of a Gaussian distribution, denoted by γ_j, are:

$$\gamma_1 = \mu \tag{4.18}$$

$$\gamma_2 = \alpha^2 \tag{4.19}$$

$$\gamma_i = 0 \text{ for } i \geq 3 \tag{4.20}$$

We may now expand the unknown non-Gaussian signal distribution f_s in terms of a known distribution with probability density function ϕ, characteristic function Φ, and standardized cumulants γ_j. The density ϕ is generally chosen to be that of the normal distribution. Using the expression of the characteristic functions for the Gaussian and non-Gaussian distributions in terms of their cumulants, we have:

$$\psi_s(t) = \exp\left[\sum_{j=1}^{\infty} \frac{(it)^j}{j!} c_j\right] = \exp\left[\sum_{j=1}^{\infty} \frac{(it)^j}{j!} \left(c_j - \gamma_j\right)\right] \Phi(t). \tag{4.21}$$

Given that $\gamma_j = c_j$ for $j = 1, 2$, and $\gamma_j = 0$ for $j > 2$, we finally have:

$$\psi_s(t) = \exp\left[\sum_{j=3}^{\infty} \frac{(it)^j}{j!} c_j\right] \Phi(t). \tag{4.22}$$

By the properties of the Fourier transform, $(it)^j \Phi(t)$ is the Fourier transform of $(-1)^j D^j \phi(x)$, where D is the differential operator with respect to x. From this, we obtain

$$f_s(x) = \exp\left[\sum_{j=3}^{\infty} c_j \frac{(-D)^j}{j!}\right] \phi(x). \tag{4.23}$$

Introducing the Hermite polynomials $H_j(\frac{x-\mu}{\alpha}) = (-1)^j \alpha^j \frac{\phi^{(j)}(x)}{\phi(x)}$, expanding the exponential and collecting terms according to the order of the derivatives, we obtain the Gram-Charlier A series, which is a second-order approximation for a distribution with mean zero and standard deviation denoted by α:

$$f_s(x) \simeq \frac{1}{\sqrt{2\pi}\alpha} \exp\left[-\frac{x^2}{2\alpha^2}\right] \left[1 + \frac{c_3}{6\alpha^3} H_3\left(\frac{x}{\alpha}\right) + \frac{c_4}{24\alpha^4} H_4\left(\frac{x}{\alpha}\right)\right], \tag{4.24}$$

where the 3rd and 4th Hermite polynomials are respectively given by $H_3(x) = x^3 - 3x$, $H_4(x) = x^4 - 6x^2 + 3$. One problem with the Gram-Charlier A series is that it is not possible to estimate the error of the expansion. For this reason, the Edgeworth expansion, formally introduced in the previous chapter, is generally preferred over the Gram-Charlier expansion. The Edgeworth expansion is based on the assumption that the unknown

signal distribution is the sum of normalized i.i.d. (non necessarily Gaussian) variables, and provides asymptotic correction terms to the Central Limit Theorem up to an order that depends on the number of moments available. When taken at the fourth-order level, we recall that the Edgeworth expansion reads as follows (see section 3.3.2):

$$f_s(x) \simeq \phi(x)\left[1+\frac{c_3}{6\alpha^3}H_3\left(\frac{x}{\alpha}\right)+\frac{c_4}{24\alpha^4}H_4\left(\frac{x}{\alpha}\right)+\frac{c_3^2}{72\alpha^6}H_6\left(\frac{x}{\alpha}\right)\right], \tag{4.25}$$

where $\phi(x)=\frac{1}{\sqrt{2\pi}\alpha}\exp\left[-\frac{x^2}{2\alpha^2}\right]$ is the density function of the Gaussian distribution, and where the 6th Hermite polynomial is defined as $H_6(x)=x^6-15x^4+45x^2-15$. We finally have $f_s(x)\simeq\phi(x)\,g(x)$ with:

$$\begin{aligned} g(x) &= \left(1+\frac{c_4}{8\alpha^4}-\frac{5c_3^2}{24\alpha^6}\right)-\frac{c_3}{2\alpha^4}x+\left(\frac{15c_3^2}{24\alpha^8}-\frac{c_4}{4\alpha^6}\right)x^2 \\ &\quad +\frac{c_3}{6\alpha^6}x^3+\left(\frac{c_4}{24\alpha^8}-\frac{5c_3^2}{24\alpha^{10}}\right)x^4+\frac{c_3^2}{72\alpha^{12}}x^6 \qquad (4.26) \\ &\equiv b_0+b_1x+b_2x^2+b_3x^3+b_4x^4+b_6x^6. \qquad (4.27) \end{aligned}$$

We see that the Edgeworth expansion involves one more Hermite polynomial with respect to the Gram-Charlier expansion while keeping the number of parameters constant. For symmetric distributions, we have $c_3=0$, and the two 4th-order expansion coincide. In general they differ by the presence of the additional term $\frac{c_3^2}{72\alpha^6}H_6\left(\frac{x}{\alpha}\right)$ in the Edgeworth expansion.

One of the problems with both of these expansions is that while the resulting approximate density does integrate to 1 (equation 3.139), it may in principal take on negative values. To see this, consider equation 4.26 in the case of a symmetric distribution ($c_3=0$) and a normalized variance $\alpha^2=1$. In this case we have:

$$g(x)=b_0+b_2x^2+b_4x^4 \tag{4.28}$$

with:

$$b_0 = 1 + \frac{c_4}{8} \tag{4.29}$$

$$b_2 = -\frac{c_4}{4} \tag{4.30}$$

$$b_4 = \frac{c_4}{24}. \tag{4.31}$$

Using the change of variable $y = x^2$, we obtain the quadratic polynomial $b_0 + b_2 y + b_4 y^2$. In case the discriminant

$$\Delta = b_2^2 - 4 b_0 b_4 = \frac{c_4^2}{16} - \frac{c_4}{6}\left(1 + \frac{c_4}{8}\right) = \frac{c_4^2}{24} - \frac{c_4}{6} = \frac{c_4}{6}\left(\frac{c_4}{4} - 1\right) \tag{4.32}$$

is strictly positive, i.e., when $c_4 > 4$, the equation

$$b_0 + b_2 y + b_4 y^2 = 0 \tag{4.33}$$

admits two real roots

$$y_1 = \frac{-b_2 - \sqrt{\Delta}}{2 b_4} = 3 - \frac{12\sqrt{\Delta}}{c_4} \tag{4.34}$$

$$y_2 = \frac{-b_2 + \sqrt{\Delta}}{2 b_4} = 3 + \frac{12\sqrt{\Delta}}{c_4} \tag{4.35}$$

and we have

$$b_0 + b_2 y + b_4 y^2 = b_4 \left(y - y_1\right)\left(y - y_2\right). \tag{4.36}$$

From this, it can be seen that the sign of the polynomial is the same as the sign of b_4, which is positive, if the discriminant Δ is strictly positive, that is if c_4 is strictly greater than 4, and if y lies outside the range $[y_1, y_2]$. In case the discriminant Δ is negative or zero, the polynomial is the same as the sign of b_4, and therefore also positive. In summary, in a case with $\alpha^2 = 1$ and $c_3 = 0$, the quantity $g\left(x\right)$ may take on negative values only if $c_4 > 4$ and these negative values will be obtained for x such that $x^2 \in \left[3 - \frac{12\sqrt{\Delta}}{c_4}, 3 + \frac{12\sqrt{\Delta}}{c_4}\right]$,

which translate into $x \in \left[\sqrt{3-\frac{12\sqrt{\Delta}}{c_4}}, \sqrt{3+\frac{12\sqrt{\Delta}}{c_4}}\right]$ or $x \in \left[-\sqrt{3+\frac{12\sqrt{\Delta}}{c_4}}, -\sqrt{3-\frac{12\sqrt{\Delta}}{c_4}}\right]$. For all tested cases in Section 4.2.1 below, we check that the chosen parameter values are such that the Edgeworth expansion always remains positive. Indeed in all cases that we test, we have $c_4 < 4$ and therefore a negative value for the discriminant Δ (given the chosen parameter values, we have $c_4 = 3$ for both the Laplace and NIG distributions, $c_4 = 3$ for the Hypersecant distribution, and $c_4 = 1.2$ for the Logistic distribution).

In general, if the parameter values are such that the discriminant Δ in equation 4.32 is positive, we can proceed by taking the square of the polynomial function g to ensure positivity of the expansion, as suggested in [81] and [82]. For example, if we want to apply the procedure to the Gram-Charlier expansion 4.24, we have:

$$f_s(x) \simeq \phi(x)\, g_{GC}(x) \tag{4.37}$$

where $\phi(x)$ is the density of a Gaussian variable with mean $\mu = 0$ and standard deviation α and where

$$\begin{aligned} g_{GC}(x) &= 1 + \frac{c_3}{6\alpha^3} H_3\left(\frac{x}{\alpha}\right) + \frac{c_4}{24\alpha^4} H_4\left(\frac{x}{\alpha}\right) && (4.38)\\ &= 1 + \frac{c_4}{8\alpha^4} - c_3\frac{x}{2\alpha^4} - c_4\frac{x^2}{4\alpha^6} + c_3\frac{x^3}{6\alpha^6} + c_4\frac{x^4}{24\alpha^8} && (4.39)\\ &\equiv a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4. && (4.40) \end{aligned}$$

Therefore:

$$\begin{aligned} g_{CG}^2(x) &= a_0^2 + a_1^2 x^2 + a_2^2 x^4 + a_3^2 x^6 + a_4^2 x^8 + 2a_0a_1x + 2a_0a_2x^2 + 2a_0a_3x^3 + 2a_0a_4x^4 \\ &\quad + 2a_1a_2x^3 + 2a_1a_3x^4 + 2a_1a_4x^5 + 2a_2a_3x^5 + 2a_2a_4x^6 + 2a_3a_4x^7 && (4.41)\\ &= a_0^2 + 2a_0a_1x + (a_1^2 + 2a_0a_2)x^2 + (2a_0a_3 + 2a_1a_2)\,x^3 + \left(a_2^2 + 2a_0a_4 + 2a_1a_3\right)x^4 \\ &\quad + (2a_1a_4 + 2a_2a_3)\,x^5 + \left(a_3^2 + 2a_2a_4\right)x^6 + 2a_3a_4x^7 + a_4^2x^8 && (4.42)\\ &\equiv c_0 + c_1x + c_2x^2 + c_3x^3 + c_4x^4 + c_5x^5 + c_6x^6 + c_7x^7 + c_8x^8. && (4.43) \end{aligned}$$

Note that a normalizing constant K needs to be introduced to ensure that the corresponding density integrates to 1:

$$\int_{-\infty}^{+\infty} \phi(x)\, h(x)\, dx = 1, \tag{4.44}$$

where $h(x) = K g_{GC}^2(x)$. We have:

$$\begin{aligned}
\int_{-\infty}^{+\infty} \phi(x)\, h(x)\, dx &= K \int_{-\infty}^{+\infty} \phi(x)\, h(x)\, dx = c_0 K + c_1 K \underbrace{\int_{-\infty}^{+\infty} x\phi(x)\, dx}_{=\mu=0} \\
&+ c_2 K \underbrace{\int_{-\infty}^{+\infty} x^2 \phi(x)\, dx}_{=\mu^2+\alpha^2=\alpha^2} + c_3 K \underbrace{\int_{-\infty}^{+\infty} x^3 \phi(x)\, dx}_{=\mu^3+3\mu\alpha^2=0} \\
&+ c_4 K \underbrace{\int_{-\infty}^{+\infty} x^4 \phi(x)\, dx}_{=\mu^4+6\mu^2\alpha^2+3\alpha^4=3\alpha^4} + c_5 K \underbrace{\int_{-\infty}^{+\infty} x^5 \phi(x)\, dx}_{=\mu^5+10\mu^3\alpha^2+15\mu\alpha^4=0} \\
&+ c_6 K \underbrace{\int_{-\infty}^{+\infty} x^6 \phi(x)\, dx}_{=\mu^6+15\mu^4\alpha^2+45\mu^2\alpha^4+15\alpha^6=15\alpha^6} \\
&+ c_7 K \underbrace{\int_{-\infty}^{+\infty} x^7 \phi(x)\, dx}_{=\mu^7+21\mu^5\alpha^2+105\mu^3\alpha^4+105\mu\alpha^6=0} \\
&+ c_8 K \underbrace{\int_{-\infty}^{+\infty} x^8 \phi(x)\, dx}_{=\mu^8+28\mu^6\alpha^2+210\mu^4\alpha^4+420\mu^2\alpha^6+105\alpha^8=105\alpha^8}
\end{aligned} \tag{4.45}$$

So finally, the constant K is given by:

$$K = \left(c_0 + c_2\alpha^2 + 3c_4\alpha^4 + 15c_6\alpha^6 + 105c_8\alpha^8\right)^{-1}. \tag{4.46}$$

4.1.2 Maximum Likelihood Estimators for the Cumulants of the SGWB Signal

We consider here a stochastic background gravitational wave signal having an unknown distribution with mean zero, variance denoted by α^2 or equivalently c_2 and third- and fourth-order cumulants denoted by c_3 and c_4, respectively. As before, the estimation procedure is based upon the following approximation for the likelihood ratio:

$$\Lambda_{ML}^{NG} = \frac{\max\limits_{\alpha,\sigma_1,\sigma_2,c_3,c_4} \int f_s|_{X=1}(s)\, f_n|_{X=1}(h-s)\, ds}{\max\limits_{\sigma_1,\sigma_2} f_n|_{X=0}(h)}. \tag{4.47}$$

Here we use the subscript NG in Λ_{ML}^{NG} to emphasize that the signal distribution is a priori non-Gaussian. Maintaining on the other hand the Gaussian assumption for the noise distribution, we obtain after straightforward manipulations the following expression for the likelihood ratio for the non-Gaussian signal distribution:

$$\Lambda_{ML}^{NG} = \max_{\alpha,\sigma_1,\sigma_2,c_3,c_4} \prod_{t=1}^{T} G_t \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(s_t-\mu_t)^2\right] g(s_t)\, ds_t, \tag{4.48}$$

with:

$$G_t \equiv \frac{\sigma}{\alpha}\frac{\overline{\sigma}_1\overline{\sigma}_2}{\sigma_1\sigma_2} \exp\left[-\frac{h_{1t}^2}{2\sigma_1^2} - \frac{h_{2t}^2}{2\sigma_2^2} + 1\right] \exp\left[\frac{1}{2}\sigma^2\left(\frac{h_{1t}}{\sigma_1^2} + \frac{h_{2t}}{\sigma_2^2}\right)^2\right], \tag{4.49}$$

and:

$$\sigma \equiv \left(\frac{1}{\alpha^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-\frac{1}{2}} \tag{4.50a}$$

$$\overline{\sigma}_i \equiv \sqrt{\frac{1}{T}\sum_{t=1}^{T} h_{it}^2} \tag{4.50b}$$

$$\mu_t \equiv \left(\frac{h_{1t}}{\sigma_1^2} + \frac{h_{2t}}{\sigma_2^2}\right)\sigma^2 \tag{4.50c}$$

Finally, we have:

$$\Lambda_{ML} = \max_{\alpha,\sigma_1,\sigma_2,c_3,c_4} \prod_{t=1}^{T} G_t \left(I_0 + I_{1t} + I_{2t} + I_{3t} + I_{4t} + I_{6t} \right), \tag{4.51}$$

with I_0, I_{1t}, I_{2t}, I_{3t}, I_{4t}, and I_{6t} given by the following Gaussian integrals:

$$\begin{aligned}
I_0 &\equiv b_0 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2} \left(s_t - \mu_t \right)^2 \right] ds_t, && (4.52a)\\
I_{1t} &\equiv b_1 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t \exp\left[-\frac{1}{2\sigma^2} \left(s_t - \mu_t \right)^2 \right] ds_t, && (4.52b)\\
I_{2t} &\equiv b_2 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t^2 \exp\left[-\frac{1}{2\sigma^2} \left(s_t - \mu_t \right)^2 \right] ds_t, && (4.52c)\\
I_{3t} &\equiv b_3 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t^3 \exp\left[-\frac{1}{2\sigma^2} \left(s_t - \mu_t \right)^2 \right] ds_t, && (4.52d)\\
I_{4t} &\equiv b_4 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t^4 \exp\left[-\frac{1}{2\sigma^2} \left(s_t - \mu \right)^2 \right] ds_t, && (4.52e)\\
I_{6t} &\equiv b_6 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t^6 \exp\left[-\frac{1}{2\sigma^2} \left(s_t - \mu_t \right)^2 \right] ds_t. && (4.52f)
\end{aligned}$$

These integrals can be obtained from the first moments of the Gaussian distribution with mean μ_t and variance σ^2, with the following results:

$$\begin{aligned}
I_0 &= b_0 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2} \left(s_t - \mu_t \right)^2 \right] ds_t = b_0, && (4.53a)\\
I_{1t} &= b_1 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t \exp\left[-\frac{1}{2\sigma^2} \left(s_t - \mu_t \right)^2 \right] ds_t = b_1 \mu_t, && (4.53b)\\
I_{2t} &= b_2 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t^2 \exp\left[-\frac{1}{2\sigma^2} \left(s_t - \mu_t \right)^2 \right] ds_t = b_2 \left(\mu_t^2 + \sigma^2 \right), && (4.53c)\\
I_{3t} &= b_3 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t^3 \exp\left[-\frac{1}{2\sigma^2} \left(s_t - \mu_t \right)^2 \right] ds_t = b_3 \left(\mu_t^3 + 3\mu_t\sigma^2 \right), && (4.53d)\\
I_{4t} &= b_4 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t^4 \exp\left[-\frac{1}{2\sigma^2} \left(s_t - \mu \right)^2 \right] ds_t = b_4 \left(\mu_t^4 + 6\mu_t^2\sigma^2 + 3\sigma^4 \right), && (4.53e)\\
I_{6t} &= b_6 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t^6 \exp\left[-\frac{1}{2\sigma^2} \left(s_t - \mu_t \right)^2 \right] ds_t,\\
&= b_6 \left(\mu_t^6 + 15\mu_t^4\sigma^2 + 45\mu_t^2\sigma^4 + 15\sigma^6 \right). && (4.53f)
\end{aligned}$$

We note that when $c_3 = c_4 = 0$, then we have $I_0 = 1$, $I_1 = I_2 = I_4 = I_6 = 0$, and we recover the standard maximum likelihood statistic (see [62]):

$$\Lambda_{ML}^{G} = \max_{\alpha,\sigma_1,\sigma_2} \prod_{t=1}^{T} \frac{\sigma}{\alpha} \frac{\overline{\sigma}_1 \overline{\sigma}_2}{\sigma_1 \sigma_2} \exp\left[-\frac{h_{1t}^2}{2\sigma_1^2} - \frac{h_{2t}^2}{2\sigma_2^2} + 1 \right] \exp\left[\frac{1}{2}\sigma^2 \left(\frac{h_{1t}}{\sigma_1^2} + \frac{h_{2t}}{\sigma_2^2} \right)^2 \right]. \qquad (4.54)$$

As recalled in the previous chapter, the likelihood ratio can be shown in the Gaussian case to admit the following simple analytical expression (equation 3.13 in [62]): $\Lambda_{ML}^{G} = \left(1 - \frac{\overline{\alpha}^4}{\overline{\sigma}_1^2 \overline{\sigma}_2^2}\right)^{-\frac{T}{2}}$, with $\overline{\alpha} = \sqrt{\frac{1}{T}\sum_{t=1}^{T} h_{1t} h_{2t}}$. From a monotonic transformation, this detection statistic is equivalent to the standard cross-correlation statistic: $\Lambda_{CC} = \sqrt{1 - \left(\Lambda_{ML}^{G}\right)^{-\frac{2}{T}}} = \frac{\overline{\alpha}^2}{\overline{\sigma}_1 \overline{\sigma}_2}$. In general, the presence of the additional terms related to higher-order cumulants implies a correction with respect to the Gaussian case. This correction makes it impossible to obtain the maximum likelihood estimate in closed-form, but straightforward numerical procedures can be used to maximize the log-likelihood function (see next section for a numerical application).

The maximum likelihood estimator is attractive since it is well-known to enjoy a number of desirable properties, including notably consistency and asymptotic efficiency. On the other hand, we now show that one can also use a moment-based method for an analytical estimation of the higher-order cumulants, thus alleviating the need to perform numerical log-likelihood maximization. The moment-based estimate for the variance of the signal coincides with the maximum likelihood estimator, but this correspondence does not extend to higher-order moments and the analytical moment-based estimators for c_3 and c_4 do not coincide in general with the maximum likelihood estimators. In numerical examples below, we find that the estimated values are relatively close, but with a lower variance for the maximum likelihood method, thus confirming for large sample sizes the superiority (asymptotic efficiency) of the maximum likelihood estimator.

To derive the moment-based estimators, we first write:

$$\begin{aligned}
\mathbb{E}\left(\mathcal{H}_1\mathcal{H}_2\right) &= \mathbb{E}\left[\left(\mathcal{N}_1+\mathcal{S}\right)\left(\mathcal{N}_2+\mathcal{S}\right)\right] && (4.55)\\
&= \mathbb{E}\left[\mathcal{N}_1\mathcal{N}_2\right]+\mathbb{E}\left[\mathcal{N}_1\mathcal{S}\right]+\mathbb{E}\left[\mathcal{N}_2\mathcal{S}\right]+\mathbb{E}\left[\mathcal{S}^2\right] && (4.56)\\
&= \mathbb{E}\left[\mathcal{N}_1\right]\mathbb{E}\left[\mathcal{N}_2\right]+\mathbb{E}\left[\mathcal{N}_1\right]\mathbb{E}\left[\mathcal{S}\right]+\mathbb{E}\left[\mathcal{S}\right]\mathbb{E}\left[\mathcal{N}_2\right]+\mathbb{E}\left[\mathcal{S}^2\right] \text{ by independence} && (4.57)\\
&= \mathbb{E}\left[\mathcal{S}^2\right]=\alpha^2 \text{ since noise and signal distributions have zero mean.} && (4.58)
\end{aligned}$$

From this analysis, we obtain that the empirical counterpart for $\mathbb{E}\left(\mathcal{H}_1\mathcal{H}_2\right)$, namely $\frac{1}{T}\sum_{t=1}^{T} h_{1t}h_{2t}$, is a natural estimator for the quantity α^2, an estimator we may call $\widehat{\alpha}^2$. It turns out that this estimator coincides with the Gaussian maximum likelihood estimator ([62]). We also have:

$$\begin{aligned}
\mathbb{E}\left(\mathcal{H}_1\mathcal{H}_2^2\right) &= \mathbb{E}\left[\left(\mathcal{N}_1+\mathcal{S}\right)\left(\mathcal{N}_2+\mathcal{S}\right)^2\right] && (4.59)\\
&= \mathbb{E}\left[\left(\mathcal{N}_1+\mathcal{S}\right)\left(\mathcal{N}_2^2+\mathcal{S}^2+2\mathcal{N}_2\mathcal{S}\right)\right] && (4.60)\\
&= \mathbb{E}\left[\mathcal{N}_1\mathcal{N}_2^2\right]+\mathbb{E}\left[\mathcal{N}_1\mathcal{S}^2\right]+2\mathbb{E}\left[\mathcal{N}_1\mathcal{N}_2\mathcal{S}\right]\\
&\quad +\mathbb{E}\left[\mathcal{S}\mathcal{N}_2^2\right]+\mathbb{E}\left[\mathcal{S}^3\right]+2\mathbb{E}\left[\mathcal{S}^2\mathcal{N}_2\right] && (4.61)\\
&= \mathbb{E}\left[\mathcal{S}^3\right]=\mu_3 \text{ since all other terms are zero.} && (4.62)
\end{aligned}$$

We thus obtain that the empirical counterpart for $\mathbb{E}\left(\mathcal{H}_1\mathcal{H}_2^2\right)$, namely $\frac{1}{T}\sum_{t=1}^{T} h_{1t}h_{2t}^2$ is a natural estimator for the quantity μ_3, an estimator we may call $\widehat{\mu}_3$. Of course, we would also have that $\mathbb{E}\left(\mathcal{H}_1^2\mathcal{H}_2\right)=\mathbb{E}\left[\mathcal{S}^3\right]=\mu_3$, so that we propose the following estimator for μ_3:

$$\widehat{\mu}_3=\frac{1}{2}\left[\frac{1}{T}\sum_{t=1}^{T} h_{1t}h_{2t}^2+\frac{1}{T}\sum_{t=1}^{T} h_{1t}^2h_{2t}\right]. \tag{4.63}$$

Finally, we have that:

$$\begin{aligned}
\mathbb{E}\left(\mathcal{H}_1^2\mathcal{H}_2^2\right) &= \mathbb{E}\left[\left(\mathcal{N}_1+\mathcal{S}\right)^2\left(\mathcal{N}_2+\mathcal{S}\right)^2\right] && (4.64)\\
&= \mathbb{E}\left[\left(\mathcal{N}_1^2+\mathcal{S}^2+2\mathcal{N}_1\mathcal{S}\right)\left(\mathcal{N}_2^2+\mathcal{S}^2+2\mathcal{N}_2\mathcal{S}\right)\right] && (4.65)\\
&= \mathbb{E}\left[\mathcal{S}^4\right]+\mathbb{E}\left[\mathcal{N}_1^2\right]\mathbb{E}\left[\mathcal{S}^2\right]+\mathbb{E}\left[\mathcal{N}_2^2\right]\mathbb{E}\left[\mathcal{S}^2\right]\\
&\quad +\mathbb{E}\left[\mathcal{N}_1^2\right]\mathbb{E}\left[\mathcal{N}_1^2\right] \text{ all other terms being zero} && (4.66)\\
&= \mu_4+\alpha^2\left(\sigma_1^2+\sigma_2^2\right)+\sigma_1^2\sigma_2^2. && (4.67)
\end{aligned}$$

If we assume that σ_1 and σ_2 are known, then we obtain the following natural estimator for μ_4:

$$\widehat{\mu}_4 = \frac{1}{T}\sum_{t=1}^{T} h_{1t}^2 h_{2t}^2 - \left(\sigma_1^2+\sigma_2^2\right)\left(\frac{1}{T}\sum_{t=1}^{T} h_{1t}h_{2t}\right) - \sigma_1^2\sigma_2^2. \tag{4.68}$$

In general, the parameters σ_1 and σ_2 are not known. The relationship

$$\mathbb{E}\left[\mathcal{H}_i^2\right] = \mathbb{E}\left[\left(\mathcal{N}_i+\mathcal{S}\right)^2\right] = \mathbb{E}\left[\mathcal{N}_i^2\right]+\mathbb{E}\left[\mathcal{S}^2\right] \tag{4.69}$$

suggests that they can be estimated as follows:

$$\widehat{\sigma}_i^2 = \overline{\sigma}_i^2 - \widehat{\alpha}^2, \tag{4.70}$$

where we recall that:

$$\overline{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T} h_{it}^2. \tag{4.71}$$

If we are instead interested in estimates for the first 4 cumulants, we have that (keeping in mind that the mean of the signal and noise distributions is zero):

$$\begin{aligned}
\widehat{\alpha}^2 &= \frac{1}{T}\sum_{t=1}^{T} h_{1t}h_{2t} && (4.72)\\
\widehat{c}_3 &= \widehat{\mu}_3 && (4.73)\\
\widehat{c}_4 &= \widehat{\mu}_4 - 3\widehat{\mu}_2 = \widehat{\mu}_4 - 3\widehat{\alpha}^4 && (4.74)
\end{aligned}$$

In the limit of vanishing noise, that is when $\sigma_1 = \sigma_2 = 0$, it is straightforward to note that:

$$\widehat{\alpha} \underset{T\to\infty}{\longrightarrow} \alpha \tag{4.75}$$

$$\widehat{c}_3 \underset{T\to\infty}{\longrightarrow} c_3 \tag{4.76}$$

$$\widehat{c}_4 \underset{T\to\infty}{\longrightarrow} c_4 \tag{4.77}$$

4.1.3 Implications for SGWB Signal Detection

For a Gaussian signal, the cross-correlation detection statistic, which can be obtained as a monotonic transformation of the likelihood ratio, is optimal in the sense of minimizing the false dismissal probability at a fixed value of the false alarm probability under restrictive assumptions ([62]). In the general non-Gaussian case, the cross-correlation detection statistic may not be optimal, and one would like to derive an optimal detection statistic when the higher order cumulants c_3 and c_4 are not zero. We address this question in chapter 5 in a more general setting involving non-Gaussian signal but also noise distributions. In what follows, we show that the application of the standard, a priori sub-optimal, cross-correlation detection statistic allows for a better estimate of the probability of a false dismissal for a given detection threshold value, or equivalently for a given probability of a false alarm, when the deviation from the Gaussian approximation is explicitly accounted for compared to a situation where the signal distribution is supposed to be Gaussian. This improvement, for which we first provide an intuitive argument before turning to a more formal analysis, is small when the signal is strongly dominated by the noise, but it is substantial when the signal and the noise are of similar magnitudes, a situation that is unlikely to be encountered in searches for stochastic gravitational wave backgrounds. Since the non-Gaussian detection methodology nests the Gaussian methodology as a specific case when the third and fourth cumulants of the signal distribution are zero, it should be noted, however, that the use of this approach can be recommended even when there is uncertainty regarding whether the signal distribution is Gaussian or

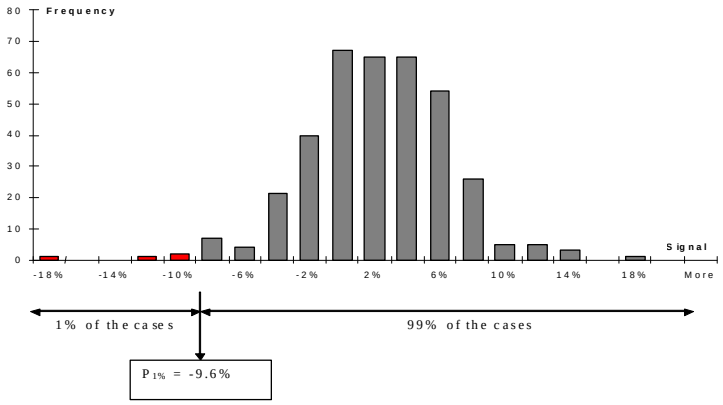

Figure 4.1: 1%-Percentile of a Generic SGWB Signal Distribution. This Figure is not meant to describe any real signal and is for illustration purposes only.

not.

An Intuitive Argument

First denote by P_x the $x-$percentile of a generic SGWB signal distribution, namely the threshold value that is such that there is only $x\%$ chances to find a value below than the threshold for the signal: $\Pr(S \leq P_x) = x\%$ (see figure 4.1).

When the signal distribution is known, and when the distribution is tabulated, then one can easily access all required information regarding the percentiles. For example, if the signal distribution is Gaussian with standard deviation α and mean value μ (here $\mu = 0$), the $x-$percentile denoted for the Gaussian distribution by P_x^G is given by $P_x^G = \mu + z_x\alpha$, where z_x denotes the x-percentile of the standardized normal density function (for example $z_{1\%} = 2.33$). When the signal distribution is unknown, one can use an inversion of the Edgeworth expansion to obtain an approximation for quantile evaluation. This approximation, known as the Cornish-Fisher expansion [47], is used in various contexts to approximate the quantiles of a random variable based only on its first few cumulants. When the distribution of the signal is assumed to show a mild deviation from the Gaussian distribution, then the 4th-order Cornish-Fisher-transformation is a suitable approximation of the true density. Accordingly, the modified normalized x-percentile ($\tilde{z}_x$) is defined

as:

$$\tilde{z}_x = z_x + \frac{1}{6}(z_x^2 - 1)Sk + \frac{1}{24}(z_x^3 - 3z_x)Kr - \frac{1}{36}(2z_x^3 - 5z_x)Sk, \tag{4.78}$$

where $Sk = \frac{c_3}{c_2^{3/2}} = \frac{c_3}{\alpha^3}$ denotes the skewness of the distribution, and $Kr = \frac{c_4}{c_2^2} = \frac{\mu_4}{\alpha^4} - 3$ the excess kurtosis of the signal distribution. As a result, we obtain the modified Cornish-Fisher estimation for the non-Gaussian percentile as $P_x^{NG} = \mu + \tilde{z}_x\alpha$. Using the Gaussian approximation, which is recovered for $Sk = Kr = 0$, typically leads to a substantial estimation error in extreme percentiles. We now consider a non-Gaussian distribution, and assume that its $x-$percentile is exactly given by the Cornish-Fisher expansion $P_x^{NG} = \mu + \tilde{z}_x\alpha$.[2] Then we consider a standard approach where a Gaussian approximation is used, with the constraint that the predicted Gaussian $x-$percentile should match the observed percentile of the signal distribution. To achieve this, a modified value $\widetilde{\alpha}_x$ (which is a function of the percentile x) for the standard deviation is used, so that the Gaussian percentile coincides with the actual non-Gaussian percentile. In other words, $\widetilde{\alpha}_x$ is the solution to $\mu + z_x\widetilde{\alpha}_x = \mu + \tilde{z}_x\alpha$, from which we obtain:

$$\widetilde{\alpha}_x = \alpha\frac{\tilde{z}_x}{z_x} = \alpha + \frac{1}{6}(z_x - \frac{1}{z_x})\frac{c_3}{\alpha^2} + \frac{1}{24}(z_x^2 - 3)\frac{c_4}{\alpha^3} - \frac{1}{36}(2z_x^2 - 5)\frac{c_3}{\alpha^2}. \tag{4.79}$$

For a non-Gaussian distribution with tails fatter than the Gaussian tails, we have $\tilde{z}_x > z_x$, for sufficiently small x values, which implies $\widetilde{\alpha}_x > \alpha$. In a nutshell, the intuition is that one should use an augmented variance when trying to generate with a Gaussian distribution the same percentiles as with a non-Gaussian distribution with fatter tails. Intuitively this analysis translates into an improved detection threshold minimizing the false dismissal rate for a fixed false alarm rate:

$$\rho_{NG}^T = \frac{\widetilde{\alpha}^2\sqrt{N}}{\sigma_1\sigma_2}, \tag{4.80}$$

[2]This is equivalent to assuming that cumulants of order strictly higher than 4 are all zero for the distribution under consideration.

which compares to the Gaussian threshold :

$$\rho_G^T = \frac{\alpha^2\sqrt{N}}{\sigma_1\sigma_2} \tag{4.81}$$

where we have $\rho_{NG}^T < \rho_G$, with a difference disappearing for $c_3 = c_4 = 0$.

A Formal Argument

We now turn to a more formal argument based on the analysis of the asymptotic distribution of the detection statistic in the Gaussian versus non-Gaussian case. We first consider the detection statistic DS given by:

$$DS = \sum_{t=1}^{T} \mathcal{H}_{1t}\mathcal{H}_{2t}, \tag{4.82}$$

where $\mathcal{H}_{it} = \mathcal{N}_{it} + \mathcal{S}_t$, and where $\mathcal{N}_{it}$ and $\mathcal{S}_t$, for $1 \leq t \leq T$, are T independent copies of the random variables $\mathcal{N}_i$ and $\mathcal{S}$, respectively. In chapter 5, we note that this detection statistic is *different* from the optimal detection statistic for normally-distributed signal and noise distributions, which displays a correction involving estimates for the variance of the detector noise distributions (see equation 5.7 in the general case or equation 5.10 when detector sensitivities are known). The developments below can be regarded as a preliminary analysis of problem, which contains explicit results regarding the distribution of the cross-correlation product of the measurement outputs $\sum_{t=1}^{T} \mathcal{H}_{1t}\mathcal{H}_{2t}$ that will be used in chapter 5.

We have:

$$DS = \sum_{t=1}^{T} \mathcal{N}_{1t}\mathcal{N}_{2t} + \sum_{t=1}^{T} \mathcal{N}_{1t}\mathcal{S}_t + \sum_{t=1}^{T} \mathcal{N}_{2t}\mathcal{S}_t + \sum_{t=1}^{T} \mathcal{S}_t^2 \tag{4.83}$$

A signal is presumed to be detected when the detection statistic DS exceeds a given detection threshold DT:

$$DS > DT \tag{4.84}$$

We typically select the detection threshold DT such that $pfa = x\%$, where pfa is the

probability of a false alarm given by:

$$pfa = \Pr\left(DS > DT \middle| \mathcal{H}_{it} = \mathcal{N}_{it}\right), \tag{4.85}$$

and where $x\%$ is a given confidence level. Obviously, we have that pfa is independent of the signal distribution, so we turn to the more interesting term, which is the probability of a false dismissal pfd given by :

$$pfd = \Pr\left(DS < DT \middle| \mathcal{H}_{it} = \mathcal{N}_{it} + \mathcal{S}_t\right). \tag{4.86}$$

By the strong law of large number, we have that (where $a.s.$ stands for *almost surely,* taken to mean that the result holds true with probability 1):

$$\sum_{t=1}^{T} \mathcal{N}_{1t}\mathcal{N}_{2t} \xrightarrow[T\to\infty]{a.s.} \mathbb{E}\left(\mathcal{N}_1\mathcal{N}_2\right) = \mathbb{E}\left(\mathcal{N}_1\right)\mathbb{E}\left(\mathcal{N}_2\right) = 0 \tag{4.87}$$

$$\sum_{t=1}^{T} \mathcal{N}_{1t}\mathcal{S} \xrightarrow[T\to\infty]{a.s.} \mathbb{E}\left(\mathcal{N}_1\mathcal{S}\right) = \mathbb{E}\left(\mathcal{N}_1\right)\mathbb{E}\left(\mathcal{S}\right) = 0 \tag{4.88}$$

$$\sum_{t=1}^{T} \mathcal{N}_{2t}\mathcal{S} \xrightarrow[T\to\infty]{a.s.} \mathbb{E}\left(\mathcal{N}_2\mathcal{S}\right) = \mathbb{E}\left(\mathcal{N}_2\right)\mathbb{E}\left(\mathcal{S}\right) = 0 \tag{4.89}$$

For finite observation times, the contribution of these three terms to the variance of the detection statistic will not vanish, and the variance of the *exact* detection statistic $\sum_{t=1}^{T} \mathcal{N}_{1t}\mathcal{N}_{2t} + \sum_{t=1}^{T} \mathcal{N}_{1t}\mathcal{S}_t + \sum_{t=1}^{T} \mathcal{N}_{2t}\mathcal{S}_t + \sum_{t=1}^{T} \mathcal{S}_t^2$ will contain contributions from the 4 terms. In what follows, we first assume that the noise is small and focus on the following approximation when the signal is present:

$$DS = \sum_{t=1}^{T} \mathcal{H}_{1t}\mathcal{H}_{2t} \simeq \sum_{t=1}^{T} \mathcal{S}_t^2. \tag{4.90}$$

In this context, we have:

$$pfd = \Pr\left(DS < DT \middle| \mathcal{H}_{it} = \mathcal{N}_{it} + \mathcal{S}_t\right) \tag{4.91}$$

$$\simeq \Pr\left(\sum_{t=1}^{T} \mathcal{S}_t^2 < DT\right). \tag{4.92}$$

When the signal is Gaussian, $\sum_{t=1}^{T} \mathcal{S}_t^2$ follows, by definition, a chi-squared distribution with T degrees of freedom.[3] On the other hand, when the signal is not Gaussian, it is not obvious to see what the distribution of the approximate detection statistic $\sum_{t=1}^{T} \mathcal{S}_t^2$ is *for a finite* T, except for very particular choices of non-Gaussian distributions. In principle, one could use an Edgeworth expansion in order to approximate the distribution of the detection statistic for each given non-Gaussian signal distribution. Fortunately, a central limit theorem exists for the sample variance, which allows us to obtain the *asymptotic* distribution of the detection statistic as T grows to infinity for *any* underlying non-Gaussian signal distribution.

Formally, let $S_1, S_2, ..., S_T$ be T i.i.d. copies of the SBGW signal, each of them with mean 0, variance $c_2 \equiv \alpha^2$, and third and fourth-order cumulants c_3 and c_4. Then, it can be shown that:

$$\Pr\left(\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} \mathcal{S}_t^2 - \alpha^2\right) < x\right) \underset{T\to\infty}{\longrightarrow} \Pr\left(U < x\right), \tag{4.93}$$

where U is a Gaussian distribution with mean zero and variance $\sigma_U^2 = c_4 + 2\alpha^4$. The proof for this result is straightforward and follows from applying the central limit theorem to squared signal distributions S^2.

In other words, we obtain that the detection statistic asymptotically converges to a Gaussian distribution, with a variance given by a function of the second and fourth cumulants. More precisely, we have $\sum_{t=1}^{T} \mathcal{S}_t^2 = T\widehat{V}_T \underset{T\to\infty}{\sim} \mathcal{N}\left(T\alpha^2, T\left(c_4 + 2\alpha^4\right)\right)$. Note that

[3] When $T \to \infty$, we know that the chi-squared distribution with T degrees of freedom converges towards a Gaussian distribution. For practical purposes, for $T > 50$, the distribution is sufficiently close to a normal distribution for the difference to be ignored [32].

the *approximate* detection statistic $\sum_{t=1}^{T} \mathcal{S}_t^2$ is closely related to the *sample* variance of the signal distribution, which is denoted by $\widehat{V}_T = \frac{1}{T}\sum_{t=1}^{T} \mathcal{S}_t^2$. $\widehat{V}_T$ admits the following asymptotic distribution $\widehat{V}_T \underset{T\to\infty}{\sim} \mathcal{N}\left(\alpha^2, \frac{1}{T}\left(c_4 + 2\alpha^4\right)\right)$, which shows that it is an asymptotically unbiased estimator for the signal variance α^2. The advantage of using $\widehat{V}_T$, as opposed to $\sum_{t=1}^{T} \mathcal{S}_t^2$, as a detection statistic is that the expectation of the former random variable does not depend on T.

When the signal is normally distributed, we have that $c_4 = 0$, and therefore $\sigma_{U,G}^2 = 2\alpha^4$, which is a standard result regarding the asymptotic distribution of the sample variance in the Gaussian case. We find that the variance of the asymptotic distribution of the signal detection statistic for non-Gaussian signal distribution $\sigma_{U,NG}^2 = c_4 + 2\alpha^4$ is always greater than the variance of the Gaussian detection $\sigma_{U,G}^2 = 2\alpha^4$. In practice, the detection threshold DT is chosen with sufficiently low value to correspond to high confidence levels (that is, pfa and pfd probabilities of 5% or 10%). In this context, because of the fatter tails of the distribution of the detection statistic in the non-Gaussian case, the detection threshold corresponding to a given pfd will be lower in the non-Gaussian case (see Figure 4.2). One implication of these findings is that if an observer wrongly uses the assumption that the signal is Gaussian ($c_4 = 0$) while the signal is truly non-Gaussian ($c_4 > 0$), then for a given confidence level, the observer using a non-Gaussian methodology will be using a lower detection threshold, which in turn will allow for the detection of fainter signals. In chapter 5, we apply this analysis to the optimal detection statistic in a setting with with known detector sensitivities.

In the realistic case when the noise is present and the signal is small compared to the noise, we have to account for the (dominating) contribution of the noise to the variance of the detection statistic. In this case, it can be shown that the detection statistic $\widehat{V}_T$ is asymptotically normally distributed, with mean α^2 and variance given by $\frac{1}{T}\left(\sigma_1^2\sigma_2^2 + \sigma_1^2\alpha^2 + \sigma_2^2\alpha^2 + 2\alpha^4 + c_4\right)$. We therefore find that the variance of the detection statistic when the non-Gaussianity of the signal is taken into account ($c_4 > 0$) is always

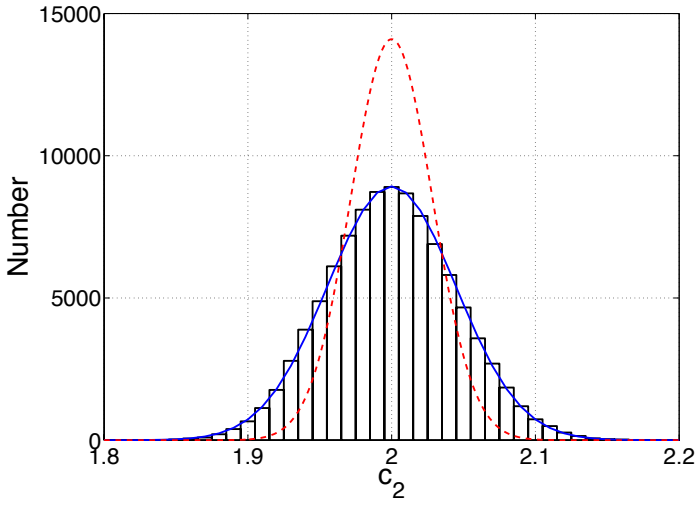

Figure 4.2: Histogram of the estimator for c_2 over 10^5 realizations for the Laplace distribution and for a number of data point $T = 10^4$. We assume here that the variance of the signal is 1 and the fourth cumulant is 3. The continuous line shows the non-Gaussian statistic density (a Gaussian centered around 2 with variance $\sigma^2_{U,NG} = c_4 + 2\alpha^4$ and the dashed line the Gaussian statistic density (a Gaussian centered in 2 with variance $\sigma^2_{U,G} = 2c_4$).

strictly greater than when the signal is assumed to be Gaussian ($c_4 = 0$). While this correction leads in principle to a sensitivity gain as explained before, the magnitude of the gain is expected to be small if the noise is several orders of magnitude larger than the signal, that is when $\frac{1}{T}\left(\sigma_1^2\sigma_2^2 + \sigma_1^2\alpha^2 + \sigma_2^2\alpha^2 + 2\alpha^4 + c_4\right) \simeq \frac{\sigma_1^2\sigma_2^2}{T}$. In chapter 5 we revisit this question in a more general context involving non-Gaussian distributions not only at the level of the SGWB signal but also at the detector noise level.

4.2 Numerical Illustrations

We now present a series of numerical illustrations showing how the methodology introduced in this chapter can be applied to estimate not only the variance but also the fourth cumulant of the signal distribution.

PDF name	Probability density	Cumulants of order j
Laplace	$\frac{a}{2}\exp\left(-a\left\|x\right\|\right)$	$c_{2j}=\frac{2(2j-1)!}{a^{2j}}$
Hypersecant	$\frac{1}{2a}\sec h\left(\frac{\pi}{2a}x\right)$	$c_{2j}=(-1)^{j+1}\left(2^{2j}-1\right)2^{2j-1}a^{2j}\frac{B_{2j}}{j}$
Logistic	$\frac{\exp\left(-\frac{x}{a}\right)}{a\left(1+\exp\left(-\frac{x}{a}\right)\right)^2}$	$c_{2j}=(-1)^{j-1}\frac{(2a\pi)^{2j}}{2j}B_{2j}$
NIG	$\frac{\delta a\exp(\delta a)}{\pi\sqrt{\delta^2+x^2}}K_1\left(a\sqrt{\delta^2+x^2}\right)$	$c_2=\frac{\delta}{a}$; $c_4=\frac{3\delta}{a^3}$

Table 4.1: Test distributions used in this paper, their probability density and cumulants. Note that the B_j are the so-called Bernoulli numbers (B_0=1, B_1=1/2, B_2=1/6, B_3=0, B_4=-1/30). The K_1 function is the Bessel function of the first kind.

4.2.1 Edgeworth Expansions of Usual Distributions

We first consider a list of 5 standard distributions, including the Gaussian distribution as well as 4 non-Gaussian distributions, namely the Laplace, Hypersecant, Logistic and Normal Inverse Gaussian distributions. These distributions are characterized in parametric form by their densities, and they have been chosen because their cumulants can conveniently be expressed as an explicit function of the parameters of the density (see Table 4.1). It should be noted that all distributions we analyze are symmetric, which implies that the parameter $c_3 = 0$. Beside, we choose the parameter values so that all distributions have a unit variance except for the Laplace distribution where we choose . More specifically, we make the following parametric choices.

1. **Laplace distribution**, we take the parameter $a = \sqrt{2}$, so that we have $c_2 = 1$, $c_3 = 0$, $c_4 = 3$; in Table 4.2 and Figure 4.7, we also test $a = 2$, which gives $c_2 = 2$, $c_3 = 0$, $c_4 = 12$, so as to confirm that the methodology generates accurate estimates for higher values of the cumulants.

2. **Hypersecant distribution**, we have that $c_2 = a^2$, $c_3 = 0$ and $c_4 = 2a^4$; taking $a = 1$, we have that $c_2 = 1$, $c_3 = 0$ and $c_4 = 2$.

3. **Logistic distribution**, we have that $c_2 = \frac{a^2\pi^2}{3}$, $c_3 = 0$ and $c_4 = \frac{2a^4\pi^4}{15}$; taking $a = \frac{\sqrt{3}}{\pi}$, we have that $c_2 = 1$, $c_3 = 0$ and $c_4 = 18/15 = 1.2$.

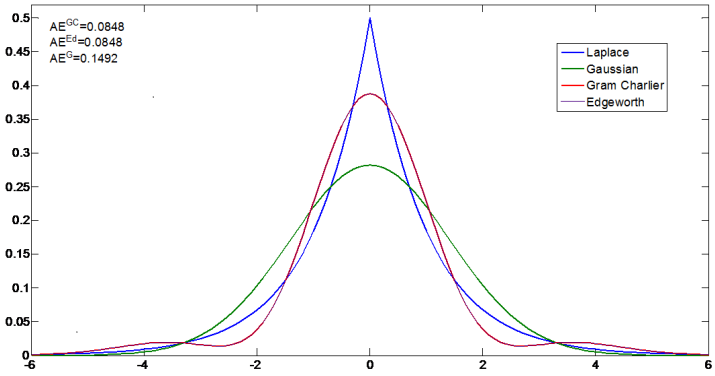

Figure 4.3: Edgeworth and Gram-Charlier approximations for the Laplace distribution.

4. **Normal Inverse Gaussian distribution**, we can take $a = \delta = 1$, so that we have $c_2 = 1$, $c_3 = 0$, $c_4 = 3$.

For each non-Gaussian distribution in the table, we plot (see Figures 4.3, 4.4, 4.5 and 4.6) on the same graph for the chosen parameter values the exact density function of the signal $f_s(x)$ as well as the approximate density function, where the approximation is given by either the Gram-Charlier or the Edgeworth expansion $f_s^E(x)$. Here the Gram-Charlier expansion exactly coincides with the Edgeworth expansion since we are only looking at symmetric distributions, so we simply focus on the latter in the following discussion. In addition to the quality of fit that can be visually assessed from the analysis of the graph, we also compute a quantitative measure of the approximation error AE as the quadratic distance between the exact and approximate density using the Edgeworth expansion:

$$AE^E = \left(\int_{-\infty}^{+\infty} \left(f_s^E(x) - f_s(x) \right)^2 dx \right)^{1/2}. \tag{4.94}$$

For comparison purposes, we also report the approximation error with the Gaussian approximation, denoted by AE^G.

The Edgeworth expansions appear to better fit the non-Gaussian density compared to the Gaussian approximation for all distributions that we consider, as can be seen from a simple visual inspection, and also more formally from the fact that approximation errors

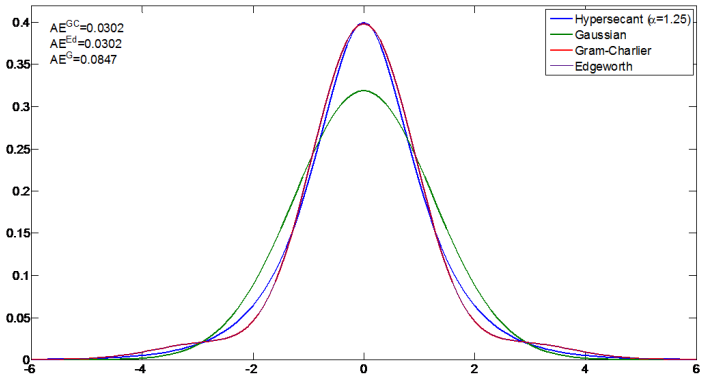

Figure 4.4: Edgeworth and Gram-Charlier approximations for the Hypersecant distribution.

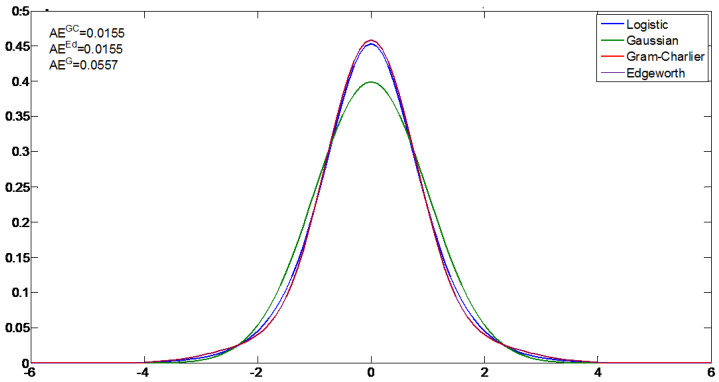

Figure 4.5: Edgeworth and Gram-Charlier approximations for the Logistic distribution.

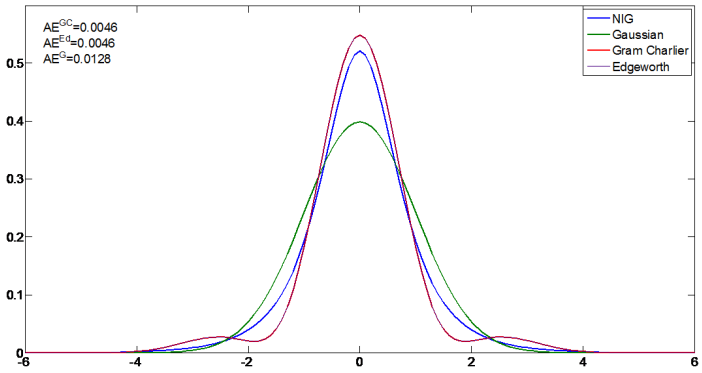

Figure 4.6: Edgeworth and Gram-Charlier approximations for the NIG distribution.

AE are between 2 and 4 times lower with the non-parametric expansions compared to what is obtained with the Gaussian approximation. The improvement stemming from using the Edgeworth expansion as opposed to the Gaussian fit ranges from 76% for the Laplace distribution ($\frac{AE^G}{AE^{Ed}} = \frac{0.1492}{0.0848} \simeq 1.76$, as shown in Figure 4.6) to 260% for the Logistic distribution ($\frac{AE^G}{AE^{Ed}} = \frac{0.0557}{0.0155} \simeq 3.6$, as shown in Figure 4.3).

4.2.2 Monte Carlo Simulations and Predictions

In order to test our new likelihood statistic, we generate fictitious data sets $h_1(t)$ and $h_2(t)$ as the output of two co-incident detectors, containing the GW signal $s(t)$, with an outcome randomly selected from the distributions presented above, and independent Gaussian noises $n_1(t)$ and $n_2(t)$. We then use the simulated data to obtain analytically the moment-based estimates for c_2 and c_4 and also to obtain numerically the maximum likelihood estimates for these parameters. The results for c_4 averaged over 10^4 trials for a number of point $T = 10^6$ and for the 4 distributions are presented in Table 4.2 and Figure 4.7. The number of points in a sample containing 1 year of data sampled at about 100 Hz will be rather of the order of 10^9 but this would have required prohibitive amount of computational resources. In order to get a realistic estimate of the performance, one should therefore divide by $\sqrt{10^3}$ the standard deviations quoted in the table. The number of trials has an effect on the average estimated value, especially when the standard deviation is large. In the limit, we expect that using an increasing number of trials would generate an estimate that would converge to the injected value in all cases. For this reason we only considered values of the ratio α^2/σ_n^2 larger than 0.03 (here we assume for simplicity that the variance of the noise is the same for both detectors, and we denote it by $\sigma_1 = \sigma_2 = \sigma_n$). With $\alpha^2/\sigma_n^2 = 0.01$, the uncertainty obtained with 10^6 points is too large and we would need to average over more than 10^4 trials to get a reasonably reliable estimate for c_4. The results we obtain are very similar for the Logistic, Hypersecant and NIG distributions. The cumulant c_2 estimated with our new statistic, which we do not report here, is exactly the same as the one derived analytically or obtained with the

Distribution		$\alpha^2\sigma_n^{-2} = 0.03$	$\alpha^2\sigma_n^{-2} = 0.05$	$\alpha^2\sigma_n^{-2} = 0.1$
Laplace	anal	12.1 (9.6)	12.0 (3.7)	12.0 (1.1)
	ML	12.1 (6.1)	11.8 (2.4)	11.3 (0.7)
Hypersecant	anal	2.0 (2.0)	2.0 (0.9)	2.0 (0.3)
	ML	2.0 (1.4)	2.0 (0.6)	2.0 (0.2)
Logistic	anal	1.2 (2.4)	1.2 (0.9)	1.2 (0.3)
	ML	1.4 (1.3)	1.2 (0.6)	1.2 (0.2)
NIG	anal	3.0 (2.4)	3.0 (0.9)	3.0 (0.3)
	ML	3.0 (1.5)	3.0 (0.6)	2.9 (0.2)

Table 4.2: Average value over 10^4 trials of c_4 estimated analytically and by likelihood maximization, for all the distributions considered in this paper, with standard deviation presented in parenthesis. The number of points in each trial is 10^6. To obtain the estimation error for 10^9 points, which corresponds to a sample of 1 year of data, one should divide the standard deviation by $\sqrt{(}10^3)$.

standard cross-correlation statistic, and is also in very good agreement (better than 1% for 1 year) with the injected value. While increasing the dimensionality of the problem by adding more parameters to estimate in principle leads to increasing the dispersion of the distribution of the estimators for the base case parameters, this result suggests that our methodology allows us to estimate c_4 accurately without any noticeable negative impact on how well estimated is the c_2 parameter, at least for the set of parameter values that we consider. Note that the ratio $\alpha^2/\sigma_n^2 = 0.01 - 0.1$ is in the range of predicted values for cosmological and astrophysical stochastic backgrounds for both Advanced LIGO and VIRGO detectors and Einstein Telescope [145]. For cosmic strings, the typical value of the energy density parameter at 100 Hz is expected to be $\Omega_{gw} \in [10^{-9}, 10^{-5}]$ which corresponds to $\alpha^2/\sigma_n^2 \in [10^{-6}, 1]$ for Advanced detectors and $\alpha^2/\sigma_n^2 \in [10^{-4}, 100]$ for Einstein Telescope; on the other hand, for compact binary mergers, $\Omega_{gw} \in [10^{-10}, 10^{-7}]$ which corresponds to $\alpha^2/\sigma_n^2 \in [10^{-7}, 0.01]$ for Advanced detectors and $\alpha^2/\sigma_n^2 \in [10^{-5}, 1]$ for Einstein Telescope [145].

We find that the fourth cumulant parameter is estimated with reasonably good precision for the cases under investigation. We also find that the maximum likelihood estimator generates an estimation error lower than that of the moment-based estimator, which is consistent with the general statement that the maximum likelihood estimator is

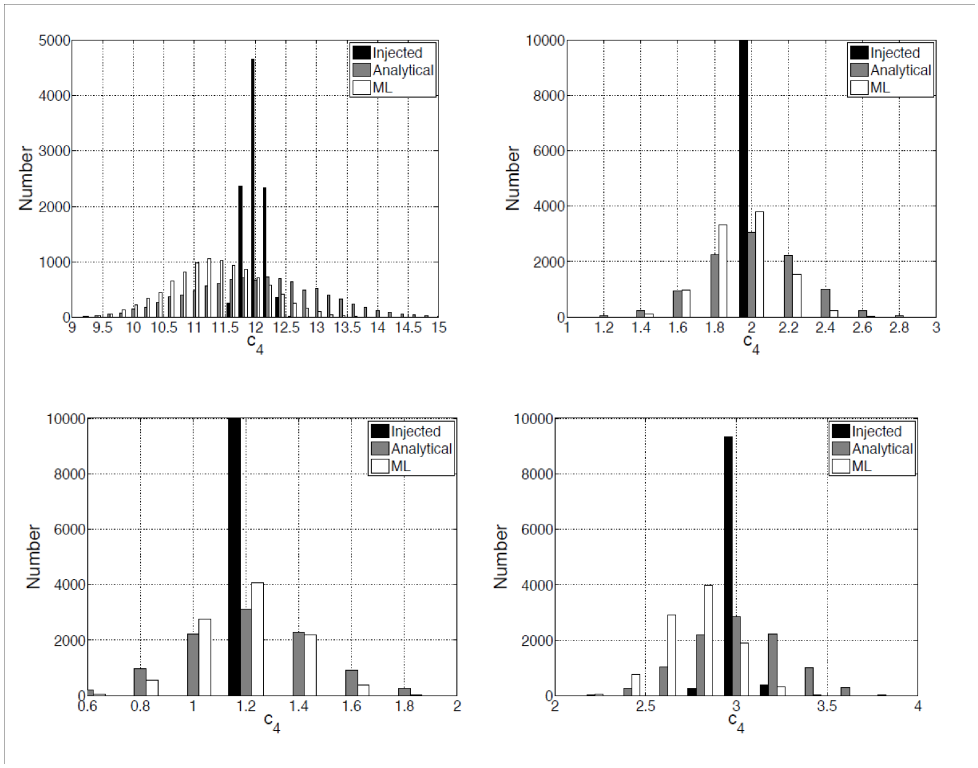

Figure 4.7: Histogram of c_4 injected (black), estimated analytically (grey) and estimated numerically by maximizing the likelihood function (white), for the distribution of Laplace (top left), Logistic (top right), hypersecant (bottom left) and NIG (bottom right). Each plot is the result of 10,000 realizations, each having 10^6 points and with $\alpha^2/\sigma_n^2 = 0.1$.

asymptotically efficient.

4.3 Extending the Approach to the Detection of Signals in the Popcorn Regime

A gravitational wave stochastic background is produced by a collection of independent gravitational wave events. In some cases, the ratio of the average time between events to the average duration of an event is small and a large number of events are taking place simultaneously. In other cases, the ratio is large and the signal received has a popcorn signature. In what has been discussed so far in this chapter, we have analyzed the first type of situations, and implicitly considered that GW events were too numerous to be individually distinguished, and yet not numerous enough for central limit theorem to give a strictly Gaussian distribution, with a deviation explicitly characterized in terms of the Edgeworth expansion. We now turn to an analysis of the second type of situations, following and generalizing an approach introduced by [62], who have focused on a model where the deviation from the Gaussian distributional assumption was understood as emanating from the presence of a resolved Gaussian signal being measured with a probability $0 < \xi \leq 1$ (the Gaussian case is recovered for $\xi = 1$). In [62], the observed distribution was assumed to be of the following form:

$$f_s(s_t) = \xi\phi(s_t) + (1-\xi)\,\delta(s_t) = \xi\frac{1}{\sqrt{2\pi}\alpha}e^{-\frac{s_t^2}{2\alpha^2}} + (1-\xi)\,\delta(s_t)\,, \tag{4.95}$$

where the parameter $0 < \xi \leq 1$ and $\delta(.)$ is the density of the Dirac distribution. This model captures a situation with a non-Gaussian signal composed of long stretches of silence which separate short bursts whose amplitudes are normally distributed, and whose durations are smaller than the detector resolution time. The parameter ξ is called in [62] the *Gaussianity parameter* of the stochastic background; it is the probability that, at any randomly chosen time, a burst is present in the detector. It can also be regarded as the

duty cycle of the background. The parameter α is the root mean square amplitude of the bursts. Since by assumption the burst events that are measured are supposed to be in small numbers, the Gaussian assumption is hard to justify and should be relaxed. We generalize this model by considering:

$$f_s(s_t) = \xi f_{NG}(s_t) + (1-\xi)\,\delta(s_t), \tag{4.96}$$

and we further assume that the unknown non-Gaussian density f_{NG} can be approximated by a 4th order Edgeworth expansion:

$$\begin{aligned} f_s(s_t) &= \xi\phi(s_t)\,g(s_t) + (1-\xi)\,\delta(s_t) && (4.97)\\ &= \xi\frac{1}{\sqrt{2\pi}\alpha}e^{-\frac{s_t^2}{2\alpha^2}}\left[1+\frac{c_3}{6\alpha^3}H_3\left(\frac{x}{\alpha}\right)+\frac{c_4}{24\alpha^4}H_4\left(\frac{x}{\alpha}\right)+\frac{c_3^2}{72\alpha^6}H_6\left(\frac{x}{\alpha}\right)\right] \\ &\quad + (1-\xi)\,\delta(s_t). && (4.98)\end{aligned}$$

Using this expression for the signal density, we obtain the following generalized form for the likelihood ratio:

$$\begin{aligned} \Lambda_{ML} &= \frac{\max\limits_{\alpha,c_3,c_4,\sigma_1,\sigma_2,\xi}\int f_s(s)\,f_n(h-s)\,ds}{\max\limits_{\sigma_1,\sigma_2} f_n} && (4.99)\\ &= \max_{\alpha,c_3,c_4,\sigma_1,\sigma_2,\xi}\prod_{t=1}^{T}\frac{\overline{\sigma}_1\overline{\sigma}_2}{\sigma_1\sigma_2}\int_{-\infty}^{+\infty} f_s(s_t)\exp\left[-\frac{(h_{1t}-s_t)^2}{2\sigma_1^2}-\frac{(h_{2t}-s_t)^2}{2\sigma_2^2}+1\right]ds_t && (4.100)\\ &= \max_{\alpha,c_3,c_4,\sigma_1,\sigma_2,\xi}\prod_{t=1}^{T}\frac{\overline{\sigma}_1\overline{\sigma}_2}{\sigma_1\sigma_2}\int_{-\infty}^{+\infty}\left[\xi f_G(s_t)\,g(s_t)+(1-\xi)\,\delta(s_t)\right] \\ &\quad \times\exp\left[-\frac{(h_{1t}-s_t)^2}{2\sigma_1^2}-\frac{(h_{2t}-s_t)^2}{2\sigma_2^2}+1\right]ds_t. && (4.101)\end{aligned}$$

After calculations similar to what has been done before for the case with $\xi = 1$, we

obtain:

$$\Lambda_{ML} = \max_{\alpha,c_3,c_4,\sigma_1,\sigma_2,\xi} \prod_{t=1}^{T} \left\{ \xi \frac{\sigma}{\alpha} \frac{\overline{\sigma}_1 \overline{\sigma}_2}{\sigma_1 \sigma_2} \exp\left[-\frac{h_{1t}^2}{2\sigma_1^2} - \frac{h_{2t}^2}{2\sigma_2^2} + 1 \right] \exp\left[\frac{1}{2}\sigma^2 \left(\frac{h_{1t}}{\sigma_1^2} + \frac{h_{2t}}{\sigma_2^2} \right)^2 \right] \right.$$
$$\left. \times (I_0 + I_{1t} + I_{2t} + I_{3t} + I_{4t} + I_{6t}) + (1-\xi) \frac{\overline{\sigma}_1 \overline{\sigma}_2}{\sigma_1 \sigma_2} \exp\left[-\frac{h_{1t}^2}{2\sigma_1^2} - \frac{h_{2t}^2}{2\sigma_2^2} + 1 \right] \right\} \quad (4.102)$$

The values of $1-\xi$, α, c_3, c_4, σ_1, σ_2, that achieve the maximum value for the likelihood function are, respectively, estimators for the probability of the presence of a (non-Gaussian) signal, the 2nd, 3rd and 4th cumulant of the signal distributions, and the variance of the noise in the two detectors. Note that if we evaluate this function at $\xi = 1$, $c_3 = c_4 = 0$, rather than maximizing over ξ, c_3, c_4, we recover the Gaussian detection statistic.

To get a sense for how strongly non-Gaussian the signal distribution can be, we have reproduced the Monte-Carlo approach described in [143] to simulate the GW signal generated by an extra-galactic population of binary black holes (BBHs). In Figure 4.8, we report the histogram obtained for the signal distribution over a time period of $T = 10^6$ seconds for an average waiting time Δt between two consecutive events taken to be equal to 1,000 seconds. As a result, we obtain an average of 1,000 simulated sources over the time interval $T = 10^6$. Assuming that there is no overlap between the sources, which is a safe assumption for an average waiting time within the range [100; 2000] discussed in [143], the fraction of the data that contains a signal is roughly $60/\Delta t$, which in this case is equal to $6 \times 10^{-3} = .6\%$. Hence the simulated data contains a large number of zeros, and the distribution in Figure 4.8 is the conditional distribution over the dates when a signal is present. We obtain for this conditional distribution a mean value of -1.2173×10^{-26} and a variance of 1.0902×10^{-45}. When looking at higher-order moments/cumulants to formally check for a non-Gaussian behavior, we find a skewness of -0.0425, signalling a quasi-symmetric distribution, but a kurtosis of 16.1461, or an excess kurtosis of $16.1461 - 3 = 13.1461$, which confirms that the data shows a strong departure from the Gaussian assumption. Overall, we obtain about 4.90×10^5 data points

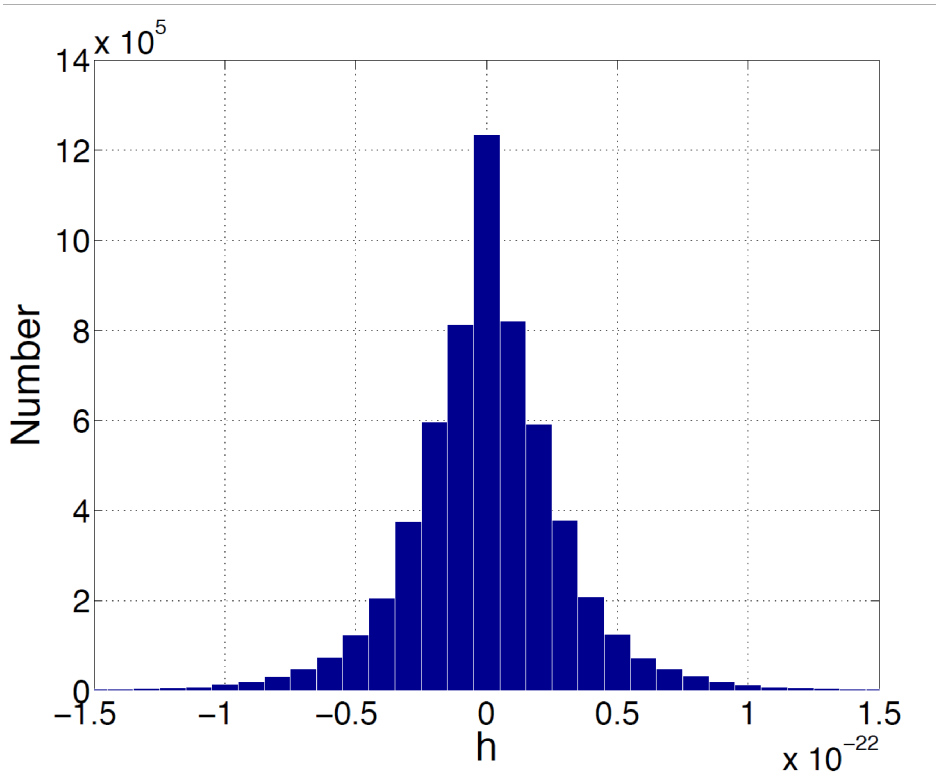

Figure 4.8: Conditional distribution of the simulated stochastic GW signal from a population of extragalactic BBHs based on a Monte-Carlo approach with simulated data for a sample of 10^6 seconds with an average waiting time observable between two consecutive BBH coalescence events of 1,000 seconds.

with a non-zero signal for a total of $2048 \times 2 \times 24 \times 3600 = 2.54 \times 10^{-29}$ observations, which corresponds to a percentage of $\xi = .6\%$. Obviously, the unconditional distribution distribution that would include the long stretches of data with no signal would be even more strongly non-Gaussian. The expectation and variance decompositions formulas can be used to obtain the mean and variance of the compound distribution $f_s(s_t) = \xi f_{NG}(s_t) + (1-\xi)\delta(s_t)$ as follows, where $X = 1$ indicates that a signal is present and $X = 0$ indicates that no signal is present:

$$\begin{aligned}
\mathbb{E}(\mathcal{S}) &= \xi \mathbb{E}(\mathcal{S}|_{X=1}) + (1-\xi)\mathbb{E}(\mathcal{S}|_{X=0}) && (4.103)\\
&= .6\% \times \left(-1.2173 \times 10^{-26}\right) = -7.3038 \times 10^{-29} && (4.104)\\
\mathbb{V}ar(\mathcal{S}) &= \mathbb{E}[\mathbb{V}ar(\mathcal{S}|X)] + \mathbb{V}ar[\mathbb{E}(\mathcal{S}|X)] && (4.105)\\
&= .6\% \times 1.0902 \times 10^{-45} + .6\% \times \left(-1.2173 \times 10^{-29} - \mathbb{E}(\mathcal{S})\right)^2 \\
&\quad + (1-.6\%) \times (0 - \mathbb{E}(\mathcal{S}))^2 && (4.106)\\
&= 6.5412 \times 10^{-48} && (4.107)
\end{aligned}$$

Similar calculation could in principle be applied to infer the skewness and the kurtosis of the compound unconditional distribution f_s from the parameter of the conditional distribution f_{NG}. However, as argued in the discussion above, we would suggest applying the Edgeworth expansion to approximate the non-Gaussian conditional distribution when the signal if present f_{NG}, as opposed to using it to approximate the full unconditional distribution f_s, which would not lend itself to an approximation given by the product of the Gaussian density and a polynomial.

5

Efficiency of the Cross-Correlation Statistic for Gravitational Wave Stochastic Backgrounds with Non-Gaussian Noise and Heterogeneous Detector Sensitivities

---

This chapter is a substantially modified version of the following paper [122]: "Efficiency of the cross-correlation statistic for gravitational wave stochastic background signals with non-Gaussian noise and heterogeneous detector sensitivities", L. Martellini and T. Regimbau, 2015, Physical Review D, 92, 10, 104025. The results presented in this paper have been found to be based on an incorrect expression for the standard detection statistic, which does not include the correction to the cross-correlation term that is required when the detectors sensitivities are assumed to be known (see for example equation 3.15 in [62] or equation 5.10 in this chapter). I am extremely grateful to Joe Romano for pointing out the problem and providing helpful feedback on the revised version of the material.

Most of the papers on SGWB detection in non-Gaussian regimes (see in particular [50], [62], [121], [157], or [168]) maintain the assumption of Gaussian noise distributions so as to better focus on the impact of deviations from normality of the signal distribution. That relatively little is known about the impact of the presence of such non-Gaussian noise distributions on the efficiency of standard methods used for the detection of SGWB signals is perhaps surprising given that there is ample evidence of strong deviations from the Gaussian assumption for noise distributions in gravitational wave detectors [13, 14], and also given that such deviations are expected to matter even more than deviations from the Gaussian assumption impacting the signal distribution since the signal is expected to be small compared to the noise in realistic detection situations. If the exact non-Gaussian nature of the detector noise is understood, it is in principle possible to repeat the construction of the optimal detection statistic using the non-Gaussian noise distribution to obtain a robust detection, and several papers have proposed a number of methodologies to deal with non-Gaussian noise for GW observations based on some specific non-Gaussian distributions such as the exponential distribution [13] or the Student's t-distribution for example [152]. Given that the actual noise distribution is a priori unknown, it would be desirable to use a nonparametric approach that would be robust to specification errors in the exact shape and nature of the deviation from the Gaussian assumption.

The main focus of this chapter is precisely to analyze the efficiency (expressed in terms of probability of a false dismissal for a given probability of a false alarm) of the standard cross-correlation (CC) statistic in situations that involve unknown forms of deviations from the Gaussian assumption. To do so we first analyze the performance of the standard cross-correlation statistic in the presence of non-Gaussian signal and noise distributions, and we derive closed-form expressions for the mean and variance of this statistic as a function of the first four cumulants of the signal and noise distributions. We find that explicitly accounting for deviations from the Gaussian assumption when applying the standard detection statistic in a specific situation with known detector sensitivities allows for a more accurate estimation of probabilities of a false dismissal for

a given probability of a false alarm with respect to a situation where the signal and noise distributions are wrongly assumed to be Gaussian. We also show how to obtain consistent estimates for the skewness and kurtosis parameters of the noise and signal distributions using a suitable extension of the likelihood function, for which we obtain an analytical expression. These results extend our results from chapter 4, where we have focused on a situation involving a non-Gaussian signal distribution, but have maintained the assumption of a Gaussian noise distribution. In addition to obtaining parameter estimates through maximum likelihood techniques, we also introduce so-called moment-based unbiased estimators given by analytical functions of the joint observations from the two detectors. Turning to a numerical analysis, we find that properly accounting for the presence of non-Gaussian distributions as opposed to wrongly assuming that higher-order cumulants of the noise distributions are zero has material implications in the implementation of standard detection procedures in that it generates substantially higher values for probabilities of false dismissal corresponding to given levels of probabilities of false alarm. The correction is found to be particularly large when detector sensitivities exhibit substantial differences, a situation that is expected to hold in early phases of development of the Advanced LIGO-Virgo detectors before they reach their design sensitivity or in joint detections from Advanced LIGO and the Einstein Telescope project [142]. In addition to outlining their implications for the performance of the standard CC detection statistic, we also discuss the implications of our results for the derivation of an optimal detection statistic in a non-Gaussian context.

The rest of this chapter is organized as follows. In section 5.1, we analyze the performance of the standard cross correlation statistic in the presence of non-Gaussian noise and signal distributions, and provide numerical estimates for the correction implied by the deviation from the Gaussian assumption. In section 5.2, we extend maximum likelihood estimation techniques to a situation involving potentially non-Gaussian signal and non-Gaussian noise distributions so as to obtain consistent estimators not only for the variance, but also for the skewness and kurtosis of the signal and noise distributions, which

are needed for correctly estimating pfa-pfd curves when using the CC detection statistic. We also discuss the derivation of an optimal detection statistic in a general setting with non-Gaussian signal and noise distributions and unknown detector sensitivities.

5.1 Performance of the Standard Cross-Correlation Statistic in the Presence of Non-Gaussian Noise (and Signal) Distributions

The standard cross-correlation detection statistic is the optimal detection statistic in the Gaussian case, but it is not necessarily optimal in the presence of non-Gaussian signal or noise distributions. Abstracting away for a moment from the problem of deriving an optimal detection statistic in a general non-Gaussian setting (see next section for a discussion of this problem), we analyze in this section how the presence of deviations from the Gaussian assumption impacts the performance of the standard detection statistic.

5.1.1 Assumptions and Notation

As in previous chapters, we consider two gravitational wave detectors. The output of each detector is a collection of dimensionless strain measurements. Suppose that N such measurements are made with each detector at regular time intervals. Denote these measurements by a $T \times 2$ matrix h with components h_t^k, where $i = 1, 2$ labels the detector, and $t = 1, 2, ..., T$ is the discrete date of measurement. We first decompose the measurement output for detector i in terms of noise versus signal, which gives when written in terms of random variables:

$$\mathcal{H}_i = \mathcal{N}_i + \mathcal{S}_i, \tag{5.1}$$

where $\mathcal{N}_i$ denotes the noise detected by the detector i and $\mathcal{S}_i$ denotes the signal detected by the detector i so that $\mathcal{H}_i$ is the total measurement for the detector i. Assuming again that the detectors are coincident and coaligned, we obtain that the signal received by

both detectors is identical:

$$\mathcal{S}_1 = \mathcal{S}_2 \equiv \mathcal{S}. \tag{5.2}$$

In terms of the realization of such random variables for either one of the two detectors, we note:

$$h_{it} = n_{it} + s_t. \tag{5.3}$$

Given that both signal and noise distributions can potentially be non-Gaussian, we denote by c_j, $j = 1, 2, 3, 4$, the first four *cumulants* of the signal distribution, and by $c_{i,j}$, $j = 1, 2, 3, 4$, the first four *cumulants* of the noise distribution for detector i, with $i = 1$ or 2. For the Gaussian distribution with mean μ and variance σ^2, we have $c_1 = \mu$, $c_2 = \sigma^2$, and $c_k = 0$ for $k > 2$. This allows us to identify deviations from the Gaussian assumption through the presence of non-zero 3rd- and 4th-order cumulants, c_3 and c_4, which are sometimes normalized so as to transform into *skewness* and *kurtosis* parameters, respectively defined as: $skw = \frac{c_3}{c_2^{3/2}}$ and $kurt = \frac{c_4}{c_2^2}$. In our application, it should finally be noted that signal and noise distributions are centered and therefore we have $c_1 = c_{1,1} = c_{2,1} = 0$. We also use the notation $c_2 = \alpha^2$, $c_{1,2} = \sigma_1^2$, and $c_{2,2} = \sigma_2^2$, where α, σ_1, and σ_2 denote the standard deviations for the signal, detector 1 and detector 2 noise distributions, respectively.

5.1.2 Evidence of non-Normality in LIGO Data

In an attempt to get a sense for how non-Gaussian the distribution of the detector noise can be, we have analyzed a small stretch of data from the first observing run (O1) collected by LIGO detector located in Livingston (Louisiana) between Oct. 25, 2015 05:28:03 UTC (GPS time 1129786100) and Oct. 25, 2015 06:01:23 UTC (GPS time 1129788100). The data is sampled at a frequency of 16 kHz, and re-sampled down by a factor of 1,024 so as to get 32,461 data points, which are displayed in Figure 5.1.

Figure 5.1 shows the presence of relatively frequent spikes of relatively large amplitude, suggesting a non-Gaussian behavior. We obtain for this sample a mean value of -

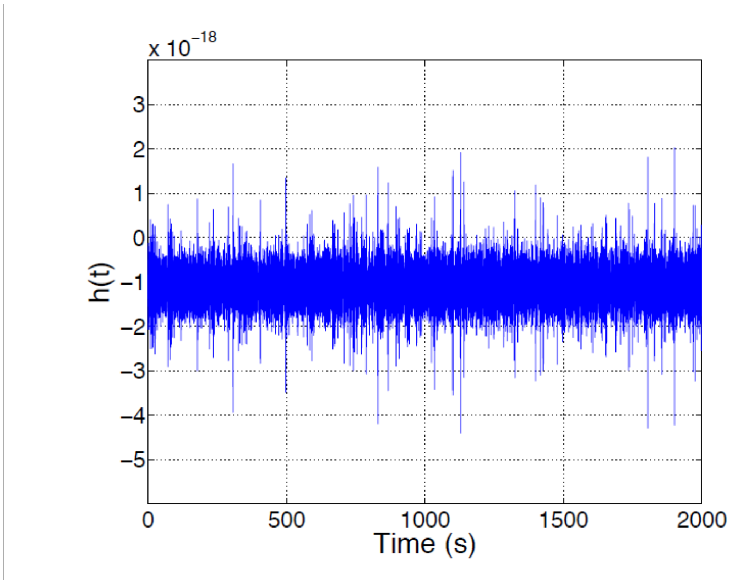

Figure 5.1: Small strech of data from the first observing run (O1) collected by LIGO detector located in Livingston (Louisiana) between Oct 25, 2015 05:28:03 UTC (GPS time 1129786100) and Oct 25, 2015 06:01:23 UTC (GPS time 1129788100).

1.09133×10^{-18} and a standard deviation of 3.46205×10^{-19}. When looking at higher-order moments/cumulants to formally check for a non-Gaussian behavior, we find a skewness of -0.00398, suggesting the presence of a quasi-symmetric distribution, but a kurtosis of 5.23, or an excess kurtosis of $5.23 - 3 = 2.23$, which confirms that the data shows a strong departure from the Gaussian assumption. Overall, this empirical check suggests that the noise distribution cannot be assumed to be Gaussian, which provides further motivation for analyzing the impact of a given deviation from the Gaussian assumption on the performance of the standard detection statistic.

5.1.3 Distribution of the Cross-Correlation Detection Statistic

To determine whether or not the data $h = (h_{1t}, h_{2t})_{t\geq 0}$ contains some desired signal, one usually compares the value of some detection statistic $DS\,(h)$ to a given threshold value DT. If $DS\,(h)$ is greater than the threshold value DT, one concludes that a signal is present and otherwise one concludes that no signal is present. A detection statistic DS is said to be optimal if it yields the smallest probability of mistakenly concluding a signal

is absent (probability of a false dismissal, or pfd) after choosing a threshold that fixes the probability for mistakenly concluding a signal is present (probability of a false alarm, or pfa).

We first recall that the standard cross correlation detection statistic DS_{cc} is given by (see for example [62]):

$$DS_{cc} = \frac{\overline{\alpha}^2}{\overline{\sigma}_1 \overline{\sigma}_2} \tag{5.4}$$

Using:

$$\overline{\alpha}^2 = \frac{1}{T}\sum_{t=1}^{T} \mathcal{H}_{1t}\mathcal{H}_{2t} \tag{5.5}$$

$$\overline{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T} \mathcal{H}_{it}^2. \tag{5.6}$$

where $\mathcal{H}_{it} = \mathcal{N}_{it} + \mathcal{S}_t$, for $i = 1, 2$, and where $\mathcal{N}_{it}$ and $\mathcal{S}_t$, for $1 \leq t \leq T$, are T independent copies of the random variables $\mathcal{N}_i$ and $\mathcal{S}$, respectively, the cross correlation detection statistic DS_{cc} can also be rewritten as:

$$DS_{cc} = \frac{\sum_{t=1}^{T} \mathcal{H}_{1t}\mathcal{H}_{2t}}{\sqrt{\sum_{t=1}^{T} \mathcal{H}_{1t}^2}\sqrt{\sum_{t=1}^{T} \mathcal{H}_{2t}^2}}, \tag{5.7}$$

This statistic can be shown to be obtained from a monotonic transformation of the Gaussian likelihood ratio ([62]), and as such is optimal (in the sense that it yields the lowest pfd for a given pfa) by the Neyman-Pearson theorem for Gaussian signal and noise distributions. Let us now specialize the analysis to a situation where detector sensitivities σ_1 and σ_2 are assumed to be known or at least sufficiently well-estimated, that is with with a precision that exceeds the size of the signal. In this case with known detector sensitivities (or kds in brief), the optimal detection statistic is *not*:

$$DS_{cc} = \frac{\overline{\alpha}^2}{\sigma_1\sigma_2} = \frac{\sum_{t=1}^{T} \mathcal{H}_{1t}\mathcal{H}_{2t}}{\sigma_1\sigma_2}, \tag{5.8}$$

but it is instead the following statistic, denoted by DS_{kds}, where *kds* stands for *known detector sensitivities* (see equation 3.15 in [62]):[1]

$$DS_{kds} = \overline{\alpha}^2 + \frac{1}{2}\left[\frac{\sigma_2^2}{\sigma_1^2}\left(\overline{\sigma}_1^2 - \sigma_1^2\right) + \frac{\sigma_1^2}{\sigma_2^2}\left(\overline{\sigma}_2^2 - \sigma_2^2\right)\right] \tag{5.9}$$

$$= \overline{\alpha}^2 + \frac{1}{2}\left[\frac{\sigma_2^2}{\sigma_1^2}\overline{\sigma}_1^2 - \sigma_2^2 + \frac{\sigma_1^2}{\sigma_2^2}\overline{\sigma}_2^2 - \sigma_1^2\right]. \tag{5.10}$$

Using again equations 5.5 and 5.6, we obtain:

$$DS_{kds} = \frac{1}{T}\sum_{t=1}^{T}\mathcal{H}_{1t}\mathcal{H}_{2t} + \frac{1}{2}\left[\frac{\sigma_2^2}{\sigma_1^2}\left(\frac{1}{T}\sum_{t=1}^{T}\mathcal{H}_{1t}^2 - \sigma_1^2\right) + \frac{\sigma_1^2}{\sigma_2^2}\left(\frac{1}{T}\sum_{t=1}^{T}\mathcal{H}_{2t}^2 - \sigma_2^2\right)\right] \tag{5.11}$$

$$\begin{aligned} = {} & \frac{1}{T}\left(\sum_{t=1}^{T}\mathcal{N}_{1t}\mathcal{N}_{2t} + \sum_{t=1}^{T}\mathcal{N}_{1t}\mathcal{S}_t + \sum_{t=1}^{T}\mathcal{N}_{2t}\mathcal{S}_t + \sum_{t=1}^{T}\mathcal{S}_t^2\right) \\ & + \frac{1}{2T}\left[\frac{\sigma_2^2}{\sigma_1^2}\left(\sum_{t=1}^{T}\mathcal{N}_{1t}^2 + 2\sum_{t=1}^{T}\mathcal{N}_{1t}\mathcal{S}_t + \sum_{t=1}^{T}\mathcal{S}_t^2 - T\sigma_1^2\right)\right] \\ & + \frac{1}{2T}\left[\frac{\sigma_1^2}{\sigma_2^2}\left(\sum_{t=1}^{T}\mathcal{N}_{2t}^2 + 2\sum_{t=1}^{T}\mathcal{N}_{2t}\mathcal{S}_t + \sum_{t=1}^{T}\mathcal{S}_t^2 - T\sigma_2^2\right)\right], \end{aligned} \tag{5.12}$$

or:

$$\begin{aligned} DS_{kds} = {} & -\frac{1}{2}\left(\sigma_1^2 + \sigma_2^2\right) + \left(\frac{1}{T}\sum_{t=1}^{T}\mathcal{S}_t^2\right)\left(1 + \frac{1}{2}\frac{\sigma_2^2}{\sigma_1^2} + \frac{1}{2}\frac{\sigma_1^2}{\sigma_2^2}\right) \\ & + \frac{1}{T}\sum_{t=1}^{T}\mathcal{N}_{1t}\mathcal{N}_{2t} + \frac{\sigma_2^2}{2\sigma_1^2}\left(\frac{1}{T}\sum_{t=1}^{T}\mathcal{N}_{1t}^2\right) + \frac{\sigma_1^2}{2\sigma_2^2}\left(\frac{1}{T}\sum_{t=1}^{T}\mathcal{N}_{2t}^2\right) \\ & + \left(\frac{1}{T}\sum_{t=1}^{T}\mathcal{N}_{1t}\mathcal{S}_t\right)\left(1 + \frac{\sigma_2^2}{\sigma_1^2}\right) + \left(\frac{1}{T}\sum_{t=1}^{T}\mathcal{N}_{2t}\mathcal{S}_t\right)\left(1 + \frac{\sigma_1^2}{\sigma_2^2}\right). \end{aligned} \tag{5.13}$$

[1] Note that we could normalize this detection statistic by multiplying by .5 and dividing by $\sigma_1\sigma_2$, in which case the expected value of the statistic can be shown to be given by the standard $SNR = \frac{\alpha^2}{\sigma_1\sigma_2}$ in case $\sigma_1 = \sigma_2$ (see equation 5.18 below).

5.1.4 Distribution of the Cross-Correlation Detection Statistic

A signal is presumed to be detected when the detection statistic DS_{kds} exceeds a given detection threshold DT:

$$DS_{kds} > DT. \tag{5.14}$$

We typically select the detection threshold DT such that $pfa = x\%$, where $x\%$ is a given confidence level (say 5% or 10%), and where the probability of a false alarm is given by the probability to exceed the threshold in a situation where there is no signal:

$$pfa = \Pr\left(DS_{kds} > DT \middle| \mathcal{H}_{it} = \mathcal{N}_{it}\right). \tag{5.15}$$

Obviously, we have that pfa is independent of the signal distribution. What depends on the signal distribution is the probability of a false dismissal pfd given by the probability that the detection statistic remains below the threshold even if there is a signal:

$$pfd = \Pr\left(DS_{kds} < DT \middle| \mathcal{H}_{it} = \mathcal{N}_{it} + \mathcal{S}_t\right). \tag{5.16}$$

By the central limit theorem applied to $\mathcal{S}_t^2$ and $\mathcal{N}_{it}^2$ it can be shown that the detection statistic DS_{cc} is asymptotically normally distributed whether or not the signal and noise distributions are Gaussian (see chapter 4 for more details in the case of a non-Gaussian signal). In this situation, the distribution of the CC detection statistic is fully characterized by its mean and variance, which can be explicitly obtained as explained below. Let us first consider the first moment of the detection statistic:

$$\mathbb{E}\left[DS_{kds}\right] = \mu_{kds} = -\frac{1}{2}\left(\sigma_1^2 + \sigma_2^2\right) + \alpha^2\left(1 + \frac{1}{2}\frac{\sigma_2^2}{\sigma_1^2} + \frac{1}{2}\frac{\sigma_1^2}{\sigma_2^2}\right) + \frac{\sigma_2^2}{2\sigma_1^2}\sigma_1^2 + \frac{\sigma_1^2}{2\sigma_2^2}\sigma_2^2 \tag{5.17}$$

$$= \alpha^2\left(1 + \frac{1}{2}\frac{\sigma_2^2}{\sigma_1^2} + \frac{1}{2}\frac{\sigma_1^2}{\sigma_2^2}\right). \tag{5.18}$$

We note that when detector sensitivities are identical ($\frac{\sigma_2^2}{\sigma_1^2} = 1$) the expression simplifies

to yield $\mu_{kds} = 2\alpha^2$. We also note that the expected value of the detection statistic is identical whether or not the non-Gaussian nature of the signal and noise distributions is accounted for since it only depends on the second moments of these distributions. Turning to the variance of the detection statistic, we obtain:

$$\begin{aligned}
\mathbb{V}ar\left[DS_{kds}\right] &= \sigma^2_{kds} = \frac{1}{T}\left(1+\frac{1}{2}\frac{\sigma_2^2}{\sigma_1^2}+\frac{1}{2}\frac{\sigma_1^2}{\sigma_2^2}\right)^2 \mathbb{V}ar\left(\mathcal{S}^2\right) \\
&\quad +\frac{1}{T}\mathbb{V}ar\left(\mathcal{N}_1\right)\mathbb{V}ar\left(\mathcal{N}_2\right)+\frac{1}{4T}\frac{\sigma_2^4}{\sigma_1^4}\mathbb{V}ar\left(\mathcal{N}_1^2\right)+\frac{1}{4T}\frac{\sigma_1^4}{\sigma_2^4}\mathbb{V}ar\left(\mathcal{N}_2^2\right) \\
&\quad +\frac{1}{T}\left(1+\frac{\sigma_2^2}{\sigma_1^2}\right)^2\sigma_1^2\alpha^2+\frac{1}{T}\left(1+\frac{\sigma_1^2}{\sigma_2^2}\right)^2\sigma_2^2\alpha^2 \qquad (5.19) \\
&= \frac{1}{T}\left(1+\frac{1}{2}\frac{\sigma_2^2}{\sigma_1^2}+\frac{1}{2}\frac{\sigma_1^2}{\sigma_2^2}\right)^2\left(c_4+2\alpha^4\right)+\frac{1}{T}\sigma_1^2\sigma_2^2 \\
&\quad +\frac{1}{4T}\frac{\sigma_2^4}{\sigma_1^4}\left(c_{1,4}+2\sigma_1^4\right)+\frac{1}{4T}\frac{\sigma_1^4}{\sigma_2^4}\left(c_{2,4}+2\sigma_2^4\right) \\
&\quad +\frac{1}{T}\left(1+\frac{\sigma_2^2}{\sigma_1^2}\right)^2\sigma_1^2\alpha^2+\frac{1}{T}\left(1+\frac{\sigma_1^2}{\sigma_2^2}\right)^2\sigma_2^2\alpha^2. \qquad (5.20)
\end{aligned}$$

When detector sensitivities are identical ($\frac{\sigma_2^2}{\sigma_1^2}=1$) the expression simplifies as follows:

$$\sigma^2_{kds} = \frac{4}{T}\left(c_4+\alpha^2\left(\sigma_1^2+\sigma_2^2\right)+2\alpha^4\right)+\frac{1}{T}\sigma_1^2\sigma_2^2+\frac{1}{4T}\left(c_{1,4}+c_{2,4}+2\sigma_1^4+2\sigma_2^4\right). \qquad (5.21)$$

5.1.5 Implications for the SGWB Signal Detection with the Standard CC Statistic

We argue that explicitly accounting for the non-Gaussian nature of the signal and noise distributions when applying the standard, a priori sub-optimal in the non-Gaussian case, cross-correlation detection statistic allows for more accurate estimates of the probability of a false dismissal for a given detection threshold value, or equivalently for a given probability of a false alarm. The difference between these estimates based on properly accounting for deviations from the Gaussian assumptions as opposed to wrongly assuming Gaussian distributions is expected to be small if the signal is not Gaussian while the noise distributions are Gaussian and if the signal is strongly dominated by the noise, but it may

be substantial when the noise distributions are also non-Gaussian, a situation that is likely to be encountered in searches for stochastic gravitational wave backgrounds. Since the non-Gaussian detection methodology nests the Gaussian methodology as a specific case when the third and fourth cumulants of the signal distribution are zero, in principle this approach can be used even when there is uncertainty regarding whether the distributions are Gaussian or not. In practice, using a less parsimonious model implies a loss of robustness in estimation procedures so the extended approach should be recommended only in cases when deviations from the Gaussian assumption are expected.

To see this, let us first observe that an application of the central limit theorem implies that the detection statistic asymptotically converges to a Gaussian distribution, with a mean μ_{kds} and a variance σ^2_{kds}. When the signal and noise are normally distributed, we have that $c_4 = 0, c_{1,4} = 0, c_{2,4} = 0$. We find that the variance of the asymptotic distribution of the signal detection statistic for non-Gaussian signal and/or noise distributions ($c_4 > 0, c_{1,4} > 0, c_{2,4} > 0$) is different (greater) when the deviations from the Gaussian assumption are accounted for compared to a situation where they are not accounted for, implying that pfd probabilities corresponding to given pfa values will be different (higher) when deviations from the Gaussian assumption are correctly accounted for with respect to the standard procedure that does not take into account higher-order cumulants. To compare the performance of the standard detection statistic when deviations are and are not taken into account, we use the following procedure.

- Step 1: We select a set of parameter values for the signal and noise distributions, and we obtain corresponding values for the mean and variance of the detection statistic in the presence of the signal, denoted respectively by $\mu_{kds,NG}$ and $\sigma^2_{kds,NG}$ to emphasize the fact that we explicitly account for the presence of non-Gaussian

(NG) distributions:

$$\mu_{kds,NG} = \alpha^2 \left(1 + \frac{1}{2}\frac{\sigma_2^2}{\sigma_1^2} + \frac{1}{2}\frac{\sigma_1^2}{\sigma_2^2}\right) \tag{5.22}$$

$$\begin{aligned}\sigma^2_{kds,NG} = {} & \frac{1}{T}\left(1 + \frac{1}{2}\frac{\sigma_2^2}{\sigma_1^2} + \frac{1}{2}\frac{\sigma_1^2}{\sigma_2^2}\right)^2 \left(c_4 + 2\alpha^4\right) + \frac{1}{T}\sigma_1^2\sigma_2^2 \\ & + \frac{1}{4T}\frac{\sigma_2^4}{\sigma_1^4}\left(c_{1,4} + 2\sigma_1^4\right) + \frac{1}{4T}\frac{\sigma_1^4}{\sigma_2^4}\left(c_{2,4} + 2\sigma_2^4\right) \\ & + \frac{1}{T}\left(1 + \frac{\sigma_2^2}{\sigma_1^2}\right)^2 \sigma_1^2\alpha^2 + \frac{1}{T}\left(1 + \frac{\sigma_1^2}{\sigma_2^2}\right)^2 \sigma_2^2\alpha^2. \end{aligned} \tag{5.23}$$

as well as the corresponding values for the mean and variance of the detection statistic in the absence of the signal (using $\alpha = c_4 = 0$), denoted respectively by $\mu^{ns}_{kds,NG}$ and $\sigma^{2,ns}_{kds,NG}$ (ns stands for *no signal*):

$$\mu^{ns}_{kds,NG} = 0 \tag{5.24}$$

$$\sigma^{2,ns}_{kds,NG} = \frac{1}{T}\sigma_1^2\sigma_2^2 + \frac{1}{4T}\left[\frac{\sigma_2^4}{\sigma_1^4}\left(c_{1,4} + 2\sigma_1^4\right) + \frac{\sigma_1^4}{\sigma_2^4}\left(c_{2,4} + 2\sigma_2^4\right)\right]. \tag{5.25}$$

- Step 2: We repeat the analysis for an observer who wrongly assumes that the distributions are Gaussian ($c_4 = c_{1,4} = c_{2,4} = 0$) to obtain corresponding values for the mean and variance of the detection statistic in the presence of the signal, denoted respectively by $\mu_{kds,G}$ and $\sigma^2_{kds,G}$:

$$\mu_{kds,G} = \alpha^2 \left(1 + \frac{1}{2}\frac{\sigma_2^2}{\sigma_1^2} + \frac{1}{2}\frac{\sigma_1^2}{\sigma_2^2}\right) = \mu_{kds,NG} \tag{5.26}$$

$$\begin{aligned}\sigma^2_{kds,G} = {} & \frac{2\alpha^4}{T}\left(1 + \frac{1}{2}\frac{\sigma_2^2}{\sigma_1^2} + \frac{1}{2}\frac{\sigma_1^2}{\sigma_2^2}\right)^2 + \frac{1}{T}\sigma_1^2\sigma_2^2 + \frac{1}{2T}\left(\sigma_1^4 + \sigma_2^4\right) \\ & + \frac{1}{T}\left(1 + \frac{\sigma_2^2}{\sigma_1^2}\right)^2 \sigma_1^2\alpha^2 + \frac{1}{T}\left(1 + \frac{\sigma_1^2}{\sigma_2^2}\right)^2 \sigma_2^2\alpha^2. \end{aligned} \tag{5.27}$$

as well as the corresponding values for the mean and variance of the detection statistic in the absence of the signal (using $\alpha = 0$), denoted respectively by $\mu^{ns}_{kds,G}$

and $\sigma^{2,ns}_{kds,G}$:

$$\mu^{ns}_{kds,G} = 0 \tag{5.28}$$

$$\sigma^{2,ns}_{kds,G} = \frac{1}{T}\sigma_1^2\sigma_2^2 + \frac{1}{2T}\left(\sigma_1^4 + \sigma_2^4\right) \tag{5.29}$$

- Step 3: For a given pfa value, we obtain the corresponding thresholds (which are functions of the selected pfa value) for the standard detection statistic when the deviations from the Gaussian assumption are properly taken into account versus not taken into account, denoted respectively by DT_{NG} and DT_G, using the following equations:

$$pfa = \Pr\left(DS_{kds} > DT_{NG}\left(pfa\right) \middle| \mathcal{H}_{it} = \mathcal{N}_{it}\right), \tag{5.30}$$

$$= \Pr\left(DS_{kds} > DT_{G}\left(pfa\right) \middle| \mathcal{H}_{it} = \mathcal{N}_{it}\right), \tag{5.31}$$

as well as the following Gaussian distributions for the detection statistic in case the signal is absent: $DS_{kds} \underset{T\to\infty}{\sim} \mathcal{N}\left(\mu^{ns}_{kds,NG}, \sigma^{2,ns}_{kds,NG}\right)$ (equations 5.24 and 5.25) when the deviations from the Gaussian assumption are accounted for, and $DS_{kds} \underset{T\to\infty}{\sim} \mathcal{N}\left(\mu^{ns}_{kds,G}, \sigma^{2,ns}_{kds,G}\right)$ (equations 5.28 and 5.29) when the noise distributions are improperly assumed to be Gaussian ($c_{1,4} = c_{2,4} = 0$). In the base case we take $T = 10^5$.

- Step 4: For all possible pfa values, we compute the probability of a false dismissal corresponding to the standard cross-correlation statistic using:

$$pfd_{NG} = \Pr\left(DS_{kds} < DT_{NG}\left(pfa\right) \middle| \mathcal{H}_{it} = \mathcal{N}_{it} + \mathcal{S}_t\right) \tag{5.32}$$

$$pfd_{G} = \Pr\left(DS_{kds} < DT_{G}\left(pfa\right) \middle| \mathcal{H}_{it} = \mathcal{N}_{it} + \mathcal{S}_t\right) \tag{5.33}$$

and the following Gaussian distributions for the detection statistic in case the signal is present $DS_{kds} \underset{T\to\infty}{\sim} \mathcal{N}\left(\mu_{kds,NG}, \sigma^{2}_{kds,NG}\right)$ (equations 5.22 and 5.23) when

the deviations from the Gaussian assumption are accounted for and $DS_{kds} \underset{T\to\infty}{\sim} \mathcal{N}\left(\mu_{kds,G}, \sigma^2_{kds,G}\right)$ (equations 5.26 and 5.27) when the noise and signal distributions are improperly assumed to be Gaussian.

- Step 5: We then plot on the same graphs (see Figure 5.2) pfd versus pfa curves for the cross-correlation statistic when the non-Gaussian nature of the noise distributions is properly accounted for (pfd_{NG}) and when it is not accounted for (pfd_G).

In Figure 5.2, we display probabilities of a false dismissal (pfd) as a function of probabilities of a false alarm (pfa) when deviations from normality are properly taken into account for (blue, red and green dotted lines corresponding to a variance of the noise distributions assumed to be identical for both detectors ($c_{1,4} = c_{2,4}$) and taking on the values 2, 4 and 6, respectively, as well as when they are not taken into account (blue solid line).[2] Note that the signal distribution is assumed to be Gaussian ($c_4 = 0$) in this analysis. In unreported results, we have analyzed the relative efficiency of the standard statistic in situations involving a non-Gaussian signal, and have found only very small differences with respect to the Gaussian signal case. Indeed, the signal is assumed to be small compared to the noise in realistic situations, and therefore the impact of deviations from the Gaussian assumption at the signal level will be dwarfed by the impact of deviations from the Gaussian assumption at the noise level. We also assume so far that detector sensitivities are identical so that ratio $r_{12} \equiv \frac{\sigma_1^2}{\sigma_2^2} = 1$. Note finally that the 3rd moment of the detector and noise distributions have no impact on the performance of the standard detection and we take them equal to 0 in our analysis.

These results suggest that properly taking into account the non-Gaussian nature of

[2]This figure has been obtained using analytical expressions for the mean and the variance of the detection statistics when the variance of the noise is assumed to be known for both detectors. Alternatively, the figure could be obtained with numerical simulations maintaining the assumption that the detector noise variances are known a priori. In principle, the figure could be slightly different since the simulated draws would on a finite sample have a variance that could slightly differ from the assumed known true population for the variance. The discrepancy is likely to be small, however. Indeed, it the variance of the estimator for the variance is given by $\mathbb{V}ar\left(\overline{\sigma}_i^2\right) = \frac{2(T-1)}{T^2}\sigma_i^2$. If we normalize the detector noise variance at the value 1, we see that the precision of the estimator is of the order of $\frac{2}{T}$. Therefore the numerical simulation will be in extremely close agreement with the analytically-derived figure if the number of observations T is taken at a realistically large value.

detector noise distributions leads to a more accurate estimation of the probabilities of a false dismissal as a function of the probability of a false alarm. More precisely we find that neglecting to account for the non-Gaussian nature of the noise distribution leads to optimistic estimates for the probabilities of a false dismissal. The required correction is extremely substantial when the deviation from normality as measured by the 4th-order cumulant of the detector noise distribution reaches a level of 2 or beyond. When $c_{1,4} = c_{2,4} = 2$, 4 and 6, the probabilities of a false dismissal for a $pfa = 5\%$ are for example $\simeq 25.5\%$, 36.21% and 44.30%, respectively, when the non-Gaussian nature of the noise distribution is accounted for, while this probability takes on a much lower value at 8.95% when it is not accounted for. For a $pfa = 10\%$, probabilities of a false dismissal when the non-Gaussian nature of the noise distribution is accounted for are $\simeq$ 15.40%, 23.76% and 30.29% for $c_{1,4} = c_{2,4} = 2$, 4 and 6, respectively, while this probability is merely 3.95% for the chosen parameter values when it is not accounted for. Focusing on $c_{1,4} = c_{2,4} = 2$, we find that for a $pfa = 5\%$ (respectively 10%) the corresponding pfd is underestimated by a factor of $25.5\%/8.95\% \simeq 2.85$ (respectively a factor of $15.4\%/3.95\% \simeq 3.90$) when the non-normality of the detector noise distribution is not properly accounted for.

While these results have been obtained in a simplified setting with known detector noise variances, it is expected that similar resulte will hold in the general case with unknown detector sensitivities. Obviously, the standard statistic is not expected to be optimal in the presence of non-Gaussian noise distribution, and lower pfd values would be obtained with an improved detection statistic that does explicitly accounts for the presence of these deviations from normality. We provide a discussion of this question in the next Section.

In addition to testing the impact of the 4th cumulant of the detector noise distributions on the performance of the standard detection statistic, we also let the ratio of the detector noise variance vary to see how the presence of detector with heterogeneous sensitivities might impact the performance of the standard statistic in the non-Gaussian setting. We

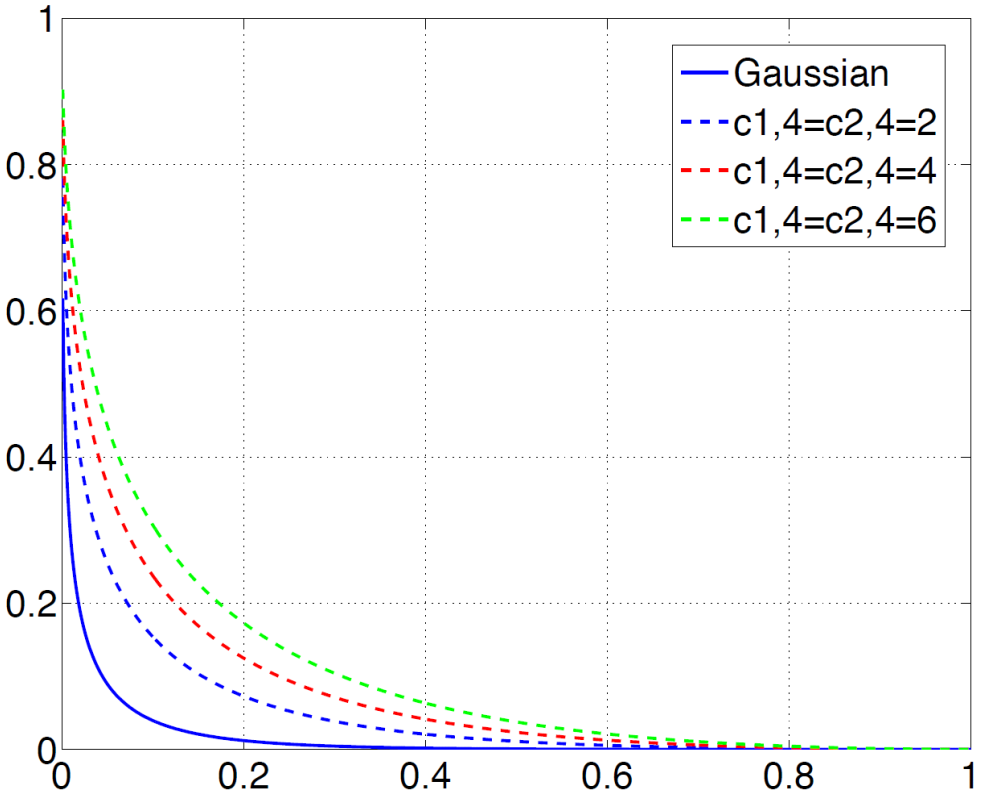

Figure 5.2: Impact of the 4th-order cumulant of detector noise distributions on the performance of the standard cross-correlation statistic. We take $T = 10^5$ and we display the probability of a false dismissal (pfd) as a function of the probability of a false alarm (pfa). We assume here that the signal is Gaussian ($c_4 = 0$) and that the 4th-order cumulant of the noise on the two detectors are identical ($c_{1,4} = c_{2,4}$). The three blue, red and green dotted lines correspond to estimated pfd when deviations from normality are properly taken into account for $c_{1,4} = c_{2,4}$ taking on the values 2, 4 and 6, respectively. The blue solid line corresponds to the pfd estimated by an observer who neglects to account for the presence of non-Gaussian noise distributions and assumes a Gaussian distribution instead. In all cases the ratio $r_{12} \equiv \dfrac{\sigma_1^2}{\sigma_2^2} = 1$ and the parameter α is chosen so that the signal to noise ratio $SNR = \sqrt{T}\dfrac{\alpha^2}{\sigma_2\sigma_2} = 2$.

expect that differences in detector sensitivities will have an impact on the performance of the standard detection statistic in non-Gaussian settings since the difference in variance of the standard detection statistic when accounting and not accounting for the non-Gaussian nature of the signal and noise distributions is increasing in the ratio $\frac{\sigma_1^2}{\sigma_2^2}$ as can be seen from the following expression:

$$\sigma^2_{kds,NG} - \sigma^2_{kds,G} = \frac{1}{T}\left(1 + \frac{1}{2}\frac{\sigma_2^2}{\sigma_1^2} + \frac{1}{2}\frac{\sigma_1^2}{\sigma_2^2}\right)^2 c_4 + \frac{1}{4T}\frac{\sigma_2^4}{\sigma_1^4}c_{1,4} + \frac{1}{4T}\frac{\sigma_1^4}{\sigma_2^4}c_{2,4}. \tag{5.34}$$

On the other hand, we have $\mu_{kds,NG} - \mu_{kds,G} = 0$ since the expected value of the detection statistic does not depend on the higher-order cumulants of the noise distribution. We note of course that $\sigma^2_{kds,NG} - \sigma^2_{kds,G} = 0$ in case the signal and noise distributions are actually Gaussian ($c_4 = 0, c_{1,4} = 0, c_{2,4} = 0$), as it should. When this ratio is equal to one (case of homogenous detectors), the expression for this difference simplifies into:

$$\sigma^2_{kds,NG} - \sigma^2_{kds,G} = \frac{4c_4}{T} + \frac{1}{4T}\left(c_{1,4} + c_{2,4}\right). \tag{5.35}$$

As a result, we expect, and confirm below, that the impact of properly accounting for given deviations from normality is even more important when detectors have largely different sensitivities.

In Figure 5.3, we plot the probability of a false dismissal (pfd) as a function of the probability of a false alarm (pfa) for different values of the ratio $r_{12} = \frac{\sigma_1^2}{\sigma_2^2}$ equal to 1, 2, 4, 10, 50 and 100 while maintaining the product $\sigma_1\sigma_2$ constant equal to 1(with no loss of generality, detectors 1 and 2 are chosen so that $r_{12} > 1$). We assume here that the signal is Gaussian ($c_4 = 0$) and that the 4th-order cumulant of the noise on the two detectors are identical and equal to 2 ($c_{1,4} = c_{2,4} = 2$). The parameter α is chosen so that the signal to noise ratio $SNR = \sqrt{T}\frac{\alpha^2}{\sigma_1\sigma_2} = 2$. We find that the probability of a false dismissal is increasing in the ratio r_{12}. For example, focusing again on $pfa = 5\%$, we find that $pfd_{NG} \simeq 25.5\%$ when $r_{12} = 1$, 30.75% when $r_{12} = 2$, 41.09% when $r_{12} = 4$, 51.14% when $r_{12} = 10$, 57.47% when $r_{12} = 50$, and 58.29% when $r_{12} = 100$. When $pfa = 10\%$,

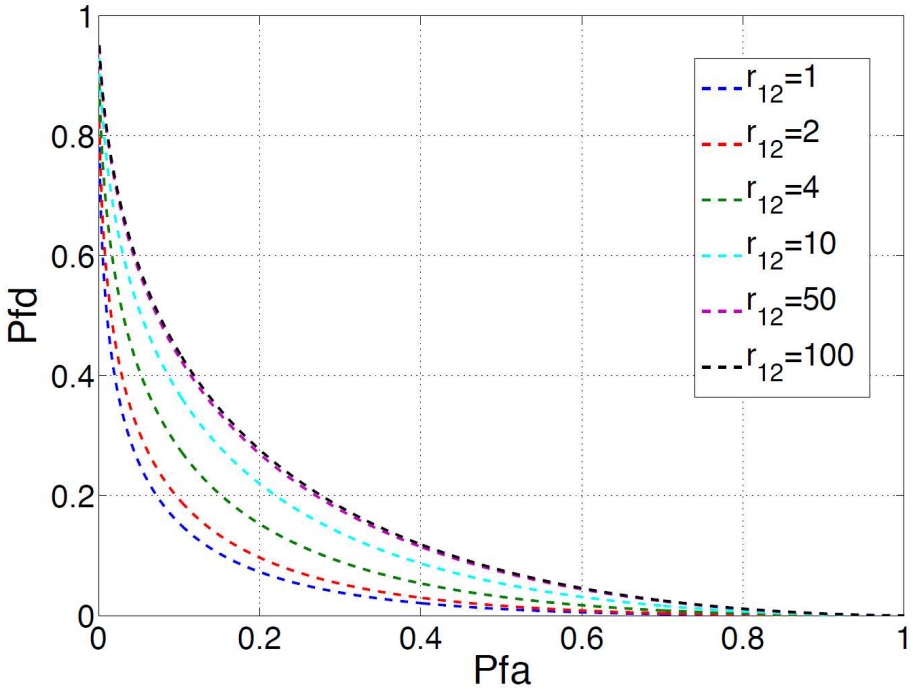

Figure 5.3: Impact of differences in detector sensitivities on the performance of the standard cross-correlation statistic in the presence of non-Gaussian noise distributions. We take $T = 10^5$ and we plot the probability of a false dismissal (pfd) as a function of the probability of a false alarm (pfa) for different values of the ratio $r_{12} = \frac{\sigma_1^2}{\sigma_2^2}$ equal to 1, 2, 4, 10, 50 and 100 while maintaining the product $\sigma_1\sigma_2$ constant equal to 1. We assume here that the signal is Gaussian ($c_4 = 0$) and that the 4th-order cumulant of the noise on the two detectors are identical and equal to 2 ($c_{1,4} = c_{2,4} = 2$). The parameter α is chosen so that the signal to noise ratio $SNR = \sqrt{T}\dfrac{\alpha^2}{\sigma_1\sigma_2} = 2$.

we have $pfd_{NG} \simeq 15.4\%$ when $r_{12} = 1$, 19.38% when $r_{12} = 2$, 27.86% when $r_{12} = 4$, 36.93% when $r_{12} = 10$, 43.07% when $r_{12} = 50$, and 43.89% when $r_{12} = 100$. In results that are not reported in Figure 5.3, we have also tested $r_{12} = 25$, 75, 200 and 500 and have obtained $pfd_{NG} \simeq 55.85\%$, 58.02%, 58.70% and 58.95% respectively for $pfa = 5\%$, and $pfd_{NG} \simeq 41.46\%$, 43.62%, 44.31% and 44.56% respectively for $pfa = 10\%$.

In a situation with heterogeneous detector sensitivities ($r_{12} = \frac{\sigma_1^2}{\sigma_2^2} > 1$) we therefore confirm that the impact of the non-Gaussian nature of the noise distributions is even more substantial than in a situation with detectors with the same sensitivities ($r_{12} = \frac{\sigma_1^2}{\sigma_2^2} = 1$). As the ratio r_{12} grows grows unboundedly large, the difference $\sigma^2_{kds,NG} - \sigma^2_{kds,G}$ goes to

Pair	σ_1^2	σ_2^2	r_{12}
AdV – aLIGO	3.6×10^{-44}	1.7×10^{-42}	48
early– middle (6 months)	3.8×10^{-41}	1.0×10^{-43}	371
middle – late (9 months)	1.5×10^{-41}	3.6×10^{-44}	402
late – design (12 months)	3.2×10^{-42}	3.6×10^{-44}	88
LIGO Red – ET-D	6.0×10^{-47}	1.9×10^{-45}	31

Table 5.1: Noise variance levels σ_1 and σ_2, and ratio $r_{12} = \frac{\sigma_1^2}{\sigma_2^2}$, calculated as $\sigma_n^2 = \int_{f_{\min}}^{f_{\max}} df S_n(f)$, where $f_{\min} = 10$ Hz and $f_{\max} = 250$ Hz is the typical frequency band used for the cross-correlation analysis, for Advanced LIGO with Advanced Virgo (aLIGO and AdV) at design sensitivities and during the early phases of development of the detectors (early, middle and late), and for Einstein Telescope (ET-D sensitivity) with LIGO Red [24], a possible Advanced LIGO sensitivity upgrade.

infinity (see equation 5.34), and the impact of the deviation from the Gaussian assumption expressed in terms of increase in pfd for a given pfa eventually reaches a fixed limit as can be seen in Figure 5.3.

This result is important because large differences in detector sensitivities are expected in practice. Realistic values for the cross-correlation between Advanced LIGO and Advanced Virgo at their nominal sensitivity can be estimated at $r_{12} = 48$, and at $r_{12} = 30$ between Einstein Telescope and LIGO Red, a possible Advanced LIGO sensitivity upgrade ([24]). We may in fact have a value $r_{12} = 400$ corresponding to the maximum expected cross-correlation between Advanced LIGO and Advanced Virgo during the early phases of the development of the detectors [2]. The projected nominal and early sensitivities, expressed in term of the square root of the power spectral density S_n, are plotted on Figure 5.4 for Advanced LIGO and Advanced Virgo [119], [166], along with the LIGO Red noise curve [24] and the proposed Einstein Telescope sensitivity ET-D [95].

The corresponding noise variances, calculated as $\sigma_n^2 = \int_{f_{\min}}^{f_{\max}} df S_n(f)$, where $f_{\min} = 10$ Hz and $f_{\max} = 250$ Hz define the typical frequency band used for the cross-correlation analysis [126], are reported in Table 5.1.5.

The green solid line in Figure 5.4 corresponding to $r_{12} = 100$ suggests that for the tested parameter values the pfd would be severely underestimated when the non-normality

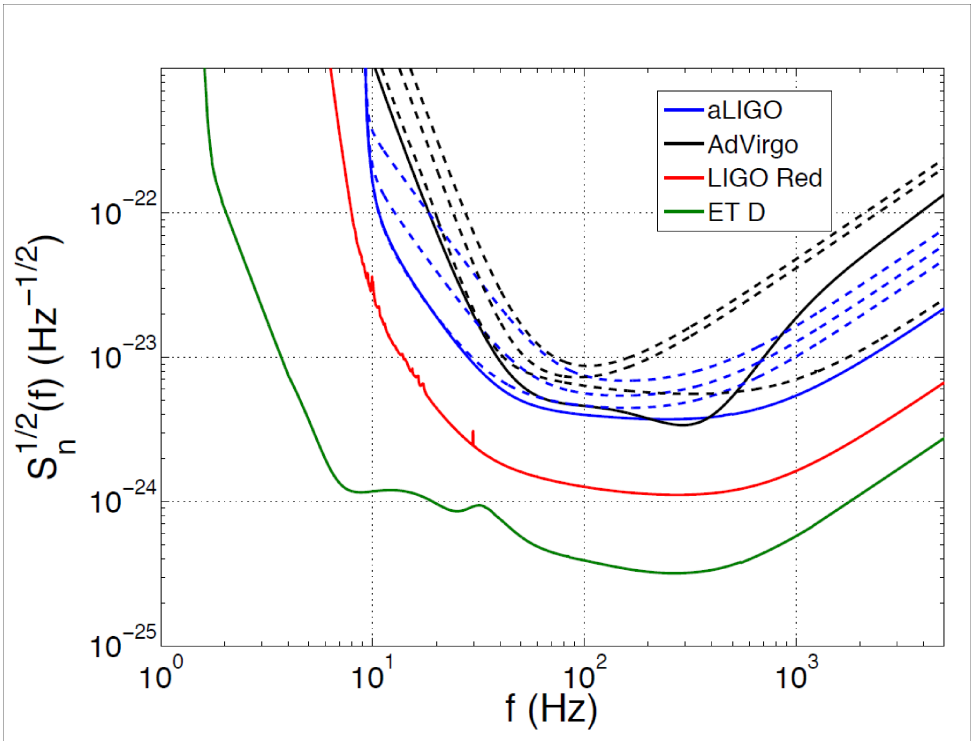

Figure 5.4: Expected sensitivity of Advanced LIGO and Advanced Virgo (blue and black continuous lines), LIGORed (dashed red line) and ET-D (green continuous line). The evolution of the sensitivity during the Advanced LIGO and Virgo early, middle and late phases are also shown in dashed blue and black lines.

of the detector noise distribution is not properly taken into account when differences in detector sensitivities reach exceedingly high levels. In realistic situations with weaker deviations from the Gaussian assumption, which would be characterized by 4th cumulant values lower than 2, our analysis suggests that ignoring the possible non-Gaussian nature of the detector noise distributions may still lead to substantially under-estimating probabilities of false dismissals in the presence of detectors with sensitivities of realistically different magnitudes.

5.2 Estimation and Detection Methods for Non-Gaussian Signal and Non-Gaussian Noise Distributions

The analysis in the previous section suggests that standard detection methodologies should be suitably corrected to account for the presence of non-Gaussian signal and noise distributions, especially when the kurtosis of the detector noise distributions take on sufficiently large values and when detector sensitivities exhibit substantial differences. It should be noted, however, that such corrections require the use of robust estimates not only for the variance, but also the skewness and kurtosis, of the signal and noise distributions. In what follows, we show how to perform maximum likelihood estimation to situations involving possibly non-Gaussian signal and noise distributions. As such, these results extend the results from the chapter 4, where we have focused on a situation involving a non-Gaussian signal distribution while maintaining the assumption of a Gaussian noise distribution.

We denote again by $f_n \equiv f_n\left(n_{1t}, n_{2t}\right)_{t=1,\ldots,T}$ the joint probability distribution for the noise in the two detectors. The standard Bayesian approach for signal detection consists in finding the value for the unknown parameters so as to minimize the false dismissal probability at a fixed value of the false alarm probability. This criterion, known as the Neyman-Pearson criterion, is uniquely defined in terms of the so-called likelihood ratio

Λ given by:

$$\Lambda = \frac{p_h|_{X=1}}{p_h|_{X=0}}, \tag{5.36}$$

where $p_h|_{X=1}$ (respectively, $p_h|_{X=0}$) is the conditional density for the measurement output if a signal is present (respectively, absent). A natural approximation of the likelihood ratio is the maximum likelihood detection statistic defined by [62]:

$$\Lambda_{ML} = \frac{\max\limits_{\alpha,\sigma_1,\sigma_2} \int f_s|_{X=1}(s)\, f_n|_{X=1}(h-s)\, ds}{\max\limits_{\sigma_1,\sigma_2} f_n|_{X=0}(h)} = \frac{N}{D}, \tag{5.37}$$

and the maximum likelihood estimators for the unknown signal and noise standard deviation parameters α, σ_1 and σ_2 are given as the corresponding likelihood maximizing quantities.

5.2.1 Full Gaussian Case

If signal and noise distributions are assumed to be Gaussian, we have:

$$f_n \equiv f_n(n_{1t}, n_{2t})_{t=1,\ldots,T} = \prod_{t=1}^{T} \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{n_{1t}^2}{2\sigma_1^2} - \frac{n_{2t}^2}{2\sigma_2^2}} \tag{5.38}$$

$$f_s \equiv f_s(s_t)_{t=1,\ldots,T} = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}\alpha} e^{-\frac{(s_t-\beta)^2}{2\alpha^2}}, \tag{5.39}$$

from which we obtain:

$$\begin{aligned}
\Lambda_{ML} &= (2\pi\overline{\sigma}_1\overline{\sigma}_2)^T \exp(T) \\
&\quad \times \max_{\alpha,\sigma_1,\sigma_2} \prod_{t=1}^{T} \int_{-\infty}^{+\infty} f_s(s_t) \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{(h_{1t}-s_t)^2}{2\sigma_1^2} - \frac{(h_{2t}-s_t)^2}{2\sigma_2^2}\right] ds_t \quad (5.40) \\
&= \max_{\alpha,\sigma_1,\sigma_2} \prod_{t=1}^{T} \frac{\overline{\sigma}_1\overline{\sigma}_2}{\sigma_1\sigma_2} \int_{-\infty}^{+\infty} f_s(s_t) \exp\left[-\frac{(h_{1t}-s_t)^2}{2\sigma_1^2} - \frac{(h_{2t}-s_t)^2}{2\sigma_2^2} + 1\right] ds_t. \quad (5.41)
\end{aligned}$$

Using:

$$f_s \equiv f_s\left(s_t\right)_{t=1,\ldots,T} = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}\alpha} e^{-\frac{s_t^2}{2\alpha^2}}, \tag{5.42}$$

we thus have:

$$\Lambda_{ML} = \max_{\alpha,\sigma_1,\sigma_2} \prod_{t=1}^{T} \frac{\overline{\sigma}_1\overline{\sigma}_2}{\sqrt{2\pi}\alpha\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left[-\frac{s_t^2}{2\alpha^2} - \frac{\left(h_{1t}-s_t\right)^2}{2\sigma_1^2} - \frac{\left(h_{2t}-s_t\right)^2}{2\sigma_2^2} + 1\right] ds_t, \tag{5.43}$$

or (see chapter 3):

$$\Lambda_{ML} = \max_{\alpha,\sigma_1,\sigma_2\geq 0} \left\{ \frac{\overline{\sigma}_1\overline{\sigma}_2}{\sqrt{\sigma_1^2\sigma_2^2+\sigma_1^2\alpha^2+\sigma_2^2\alpha^2}} \exp\left[\frac{\frac{\overline{\sigma}_1^2}{\sigma_1^4}+\frac{\overline{\sigma}_2^2}{\sigma_2^4}+\frac{2\overline{\alpha}^2}{\sigma_1^2\sigma_2^2}}{2\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}+\frac{1}{\alpha^2}\right)} - \frac{\overline{\sigma}_1^2}{2\sigma_1^2} - \frac{\overline{\sigma}_2^2}{2\sigma_2^2} + 1 \right] \right\}^T. \tag{5.44}$$

5.2.2 Gaussian Signal and Non-Gaussian Noise

In chapter 4, the Gaussian assumption was maintained for the detector noise distribution, but relaxed for the signal distribution. In what follows, we consider the opposite situation, namely a normally distributed signal, and a potentially non-Gaussian noise distribution. In other words, we assume:

$$f_s\left(s_t\right) = \frac{1}{\sqrt{2\pi}\alpha} e^{-\frac{s_t^2}{2\alpha^2}} \tag{5.45}$$

$$f_{n_i}\left(n_{it}\right) \neq \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{n_{it}^2}{2\sigma_i^2}} \text{ for } i=1,2. \tag{5.46}$$

We propose again to use a semi-parametric approach which allows one to approximate the unknown density as a transformation of a reference function (typically the Gaussian density), involving higher-order moments/cumulants of the unknown distribution. In what follows, we show that a non-parametric expansion of the unknown signal density function allows us to obtain an analytical derivation of the nearly optimal maximum likelihood detection statistic for non-Gaussian detector noise. More precisely, we want to approximate f_{n_i}, the density function of the unknown distribution of the noise dis-

tribution $\mathcal{N}_i$, as a function of the Gaussian density function $\phi_{n_i}(x)$ and a multiplicative deviation from the Gaussian density function $g_{n_i}(x)$. To achieve this end, we use again the Edgeworth expansion:

$$f_{n_i}(x) \simeq \phi_{n_i}(x)\left[1+\frac{c_{i,3}}{6\sigma_i^3}H_3\left(\frac{x}{\sigma_i}\right)+\frac{c_{i,4}}{24\sigma_i^4}H_4\left(\frac{x}{\sigma_i}\right)+\frac{c_{i,3}^2}{72\sigma_i^6}H_6\left(\frac{x}{\sigma_i}\right)\right], \tag{5.47}$$

where the 6th Hermite polynomial is defined as $H_6(x) = x^6 - 15x^4 + 45x^2 - 15$. Here the Edgeworth expansion is applied directly to the noise distribution to approximate its density through a Gaussian density and a correction term. This is different from the application of the Edgeworth expansion to the stochastic background signal from astrophysical origin, which can be regarded as the sum of overlapping signals and where the Edgeworth expansion provides a convergence rate with respect to the number of such overlapping signals. We finally have $f_{n_i}(x) \simeq \phi_{n_i}(x)\, g_{n_i}(x)$ with:

$$\phi_{n_i}(x) \equiv \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left[-\frac{x^2}{2\sigma_i^2}\right], \tag{5.48}$$

$$\begin{aligned} g_{n_i}(x) \equiv\ & \underbrace{\left(1+\frac{c_{i,4}}{8\sigma_i^4}-\frac{5c_{i,3}^2}{24\sigma_i^6}\right)}_{b_{i,0}}\underbrace{-\frac{c_{i,3}}{2\sigma_i^4}}_{b_{i,1}}x+\underbrace{\left(\frac{15c_{i,3}^2}{24\sigma_i^8}-\frac{c_{i,4}}{4\sigma_i^6}\right)}_{b_{i,2}}x^2+\underbrace{\frac{c_{i,3}}{6\sigma_i^6}}_{b_{i,3}}x^3 \\ & +\underbrace{\left(\frac{c_{i,4}}{24\sigma_i^8}-\frac{5c_{i,3}^2}{24\sigma_i^{10}}\right)}_{b_{i,4}}x^4+\underbrace{\frac{c_{i,3}^2}{72\sigma_i^6}}_{b_{i,6}}x^6. \end{aligned} \tag{5.49}$$

In this context, the likelihood detection statistic becomes:

$$\Lambda_{ML} = \frac{\max\limits_{\alpha,\sigma_1,\sigma_2,c_{1,3},c_{1,4},c_{2,3},c_{2,4}} \int f_s(s)\, f_n(h-s)\, ds}{\max\limits_{\sigma_1,\sigma_2,c_{1,3},c_{1,4},c_{2,3},c_{2,4}} f_n(h)} = \frac{N}{D} \tag{5.50}$$

The numerator N of this expression is given by:

$$\begin{aligned} N &= \max_{\alpha,\sigma_1,\sigma_2,c_{1,3},c_{1,4},c_{2,3},c_{2,4}} \prod_{t=1}^{T} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\alpha 2\pi\sigma_1\sigma_2} \exp\left[-\frac{s_t^2}{2\alpha^2} - \frac{(h_{1t}-s_t)^2}{2\sigma_1^2} - \frac{(h_{2t}-s_t)^2}{2\sigma_2^2}\right] \\ &\quad \times g_{n_1}(s_t)\, g_{n_2}(s_t)\, ds_t \qquad (5.51) \\ &= \max_{\alpha,\sigma_1,\sigma_2,c_{1,3},c_{1,4},c_{2,3},c_{2,4}} \frac{1}{(2\pi\sigma_1\sigma_2)^T} \prod_{t=1}^{T} \exp\left[-\frac{h_{1t}^2}{2\sigma_1^2} - \frac{h_{2t}^2}{2\sigma_2^2}\right] \exp\left[\frac{1}{2}\sigma^2\left(\frac{h_{1t}}{\sigma_1^2} + \frac{h_{2t}}{\sigma_2^2}\right)^2\right] \\ &\quad \times \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(s_t-\mu_t)^2\right] g_{n_1}(s_t)\, g_{n_2}(s_t)\, ds_t. \qquad (5.52) \end{aligned}$$

with:

$$\sigma = \left(\frac{1}{\alpha^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-\frac{1}{2}} \qquad (5.53)$$

$$\mu_t = \left(\frac{h_{1t}}{\sigma_1^2} + \frac{h_{2t}}{\sigma_2^2}\right)\sigma^2 \qquad (5.54)$$

Focusing for simplicity on symmetric noise distribution functions (therefore such that $c_{i,3} = 0$), we have:

$$\begin{aligned} g_{n_1}(x)\, g_{n_2}(x) &= \left(b_{1,0} + b_{1,2}x^2 + b_{1,4}x^4\right)\left(b_{2,0} + b_{2,2}x^2 + b_{2,4}x^4\right) \qquad (5.55) \\ &= \beta_0 + \beta_2 x^2 + \beta_4 x^4 + \beta_6 x^6 + \beta_8 x^8, \qquad (5.56) \end{aligned}$$

with:

$$\beta_0 = b_{1,0}b_{2,0} \qquad (5.57)$$

$$\beta_2 = b_{1,0}b_{2,2} + b_{1,2}b_{2,0} \qquad (5.58)$$

$$\beta_4 = b_{1,0}b_{2,4} + b_{1,2}b_{2,2} + b_{2,0}b_{1,4} \qquad (5.59)$$

$$\beta_6 = b_{1,4}b_{2,2} + b_{1,2}b_{2,4} \qquad (5.60)$$

$$\beta_8 = b_{1,4}b_{2,4}. \qquad (5.61)$$

So we need to compute the following integrals, which can be obtained from the first

moments of the Gaussian distribution:

$$I_0 = \beta_0 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}\left(s_t-\mu_t\right)^2\right] ds_t = \beta_0 \tag{5.62}$$

$$I_{2t} = \beta_2 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t^2 \exp\left[-\frac{1}{2\sigma^2}\left(s_t-\mu_t\right)^2\right] ds_t = \beta_2\left(\mu_t^2+\sigma^2\right) \tag{5.63}$$

$$= \beta_2\left(\left(\frac{h_{1t}}{\sigma_1^2}+\frac{h_{2t}}{\sigma_2^2}\right)^2\sigma^4+\sigma^2\right) \tag{5.64}$$

$$I_{4t} = \beta_4 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t^4 \exp\left[-\frac{1}{2\sigma^2}\left(s_t-\mu\right)^2\right] ds_t = \beta_4\left(\mu_t^4+6\mu_t^2\sigma^2+3\sigma^4\right) \tag{5.65}$$

$$I_{6t} = \beta_6 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t^6 \exp\left[-\frac{1}{2\sigma^2}\left(s_t-\mu_t\right)^2\right] ds_t \tag{5.66}$$

$$= \beta_6\left(\mu_t^6+15\mu_t^4\sigma^2+45\mu_t^2\sigma^4+15\sigma^6\right) \tag{5.67}$$

$$I_{8t} = \beta_8 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t^8 \exp\left[-\frac{1}{2\sigma^2}\left(s_t-\mu_t\right)^2\right] ds_t \tag{5.68}$$

$$= \beta_8\left(\mu_t^8+28\mu_t^6\sigma^2+210\mu_t^4\sigma^4+420\mu_t^2\sigma^6+105\sigma^8\right) \tag{5.69}$$

Finally, we have that:

$$N = \max_{\alpha,\sigma_1,\sigma_2,c_{1,4},c_{2,4}} \frac{1}{\left(2\pi\sigma_1\sigma_2\right)^T}\left\{\prod_{t=1}^{T}\frac{\sigma}{\alpha}\exp\left[-\frac{h_{1t}^2}{2\sigma_1^2}-\frac{h_{2t}^2}{2\sigma_2^2}\right]\exp\left[\frac{1}{2}\sigma^2\left(\frac{h_{1t}}{\sigma_1^2}+\frac{h_{2t}}{\sigma_2^2}\right)^2\right]\right.$$
$$\left.\times\left(I_0+I_{1t}+I_{2t}+I_{3t}+I_{4t}+I_{6t}+I_{8t}\right)\right\}. \tag{5.70}$$

We note that when $c_{1,4}=c_{2,4}=0$, that is when the third and fourth-order cumulants vanish, as is the case for a Gaussian distribution, then we have $I_0=1$, $I_1=I_2=I_4=I_6=I_{8t}=0$, and we recover the expression of the numerator for the Gaussian case:

$$N = \max_{\alpha,\sigma_1,\sigma_2} \frac{1}{\left(2\pi\sigma_1\sigma_2\right)^T}\prod_{t=1}^{T}\frac{\sigma}{\alpha}\exp\left[-\frac{h_{1t}^2}{2\sigma_1^2}-\frac{h_{2t}^2}{2\sigma_2^2}\right]\exp\left[\frac{1}{2}\sigma^2\left(\frac{h_{1t}}{\sigma_1^2}+\frac{h_{2t}}{\sigma_2^2}\right)^2\right]. \tag{5.71}$$

Similar (and simpler) expressions can be obtained for the denominator D, which relates to the situation when there is no signal. In the Gaussian case, one could obtain the following

explicit expressions for the variables involved in the maximization of the numerator N of the likelihood detection statistic [62]:

$$\alpha^2 = \widehat{\alpha}^2 \equiv \left(\overline{\alpha}^2\right)^+ \tag{5.72}$$

$$\sigma_i^2 = \widehat{\sigma}_i^2 \equiv \left(\overline{\sigma}_i^2 - \widehat{\alpha}^2\right)^+ \tag{5.73}$$

and for the variables involved in the maximization of the denominator D of the likelihood detection statistic:

$$\sigma_i^2 = \overline{\sigma}_i^2 \tag{5.74}$$

Here, the expression for the numerator of the likelihood detection statistic is more involved and it is not clear whether any explicit solutions can be obtained for the values for $\alpha, \sigma_1, \sigma_2, c_{1,4}, c_{2,4}$ that would lead to the maximum. In this context, one needs to resort to numerical optimization procedures for maximizing independently the numerator and the denominator of the likelihood statistic.

5.2.3 Non-Gaussian Signal and Noise Distributions

The methodology can also be extended to account for the presence of deviations from the Gaussian assumption for both the signal and noise distributions. To do so, we use again the Edgeworth expansion to approximate the unknown signal and noise distributions as

$f_s(x) \simeq \phi_s(x)\, g_s(x)$ and $f_{n_i}(x) \simeq \phi_{n_i}(x)\, g_{n_i}(x)$ with:

$$\phi_s(x) \equiv \frac{1}{\sqrt{2\pi}\alpha} \exp\left[-\frac{x^2}{2\alpha^2}\right] \tag{5.75}$$

$$\begin{aligned} g_s(x) \equiv & \underbrace{\left(1 + \frac{c_4}{8\alpha^4} - \frac{5c_3^2}{24\alpha^6}\right)}_{b_0} \underbrace{-\frac{c_3}{2\alpha^4}}_{b_1} x + \underbrace{\left(\frac{15c_3^2}{24\alpha^8} - \frac{c_4}{4\alpha^6}\right)}_{b_2} x^2 + \underbrace{\frac{c_3}{6\alpha^6}}_{b_3} x^3 \\ & + \underbrace{\left(\frac{c_4}{24\alpha^8} - \frac{5c_3^2}{24\alpha^{10}}\right)}_{b_4} x^4 + \underbrace{\frac{c_3^2}{72\alpha^6}}_{b_6} x^6 \end{aligned} \tag{5.76}$$

$$\phi_{n_i}(x) \equiv \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{x^2}{2\sigma_i^2}\right] \tag{5.77}$$

$$\begin{aligned} g_{n_i}(x) \equiv & \underbrace{\left(1 + \frac{c_{i,4}}{8\sigma_i^4} - \frac{5c_{i,3}^2}{24\sigma_i^6}\right)}_{b_{i,0}} \underbrace{-\frac{c_{i,3}}{2\sigma_i^4}}_{b_{i,1}} x + \underbrace{\left(\frac{15c_{i,3}^2}{24\sigma_i^8} - \frac{c_{i,4}}{4\sigma_i^6}\right)}_{b_{i,2}} x^2 + \underbrace{\frac{c_{i,3}}{6\sigma_i^6}}_{b_{i,3}} x^3 \\ & + \underbrace{\left(\frac{c_{i,4}}{24\sigma_i^8} - \frac{5c_{i,3}^2}{24\sigma_i^{10}}\right)}_{b_{i,4}} x^4 + \underbrace{\frac{c_{i,3}^2}{72\sigma_i^6}}_{b_{i,6}} x^6 \end{aligned} \tag{5.78}$$

In this context, the numerator N of the likelihood function becomes:

$$\begin{aligned} N = & \max_{\substack{\alpha, \sigma_1, \sigma_2, c_3, c_4, \\ c_{1,3}, c_{1,4}, c_{2,3}, c_{2,4}}} \frac{1}{(2\pi\sigma_1\sigma_2)^T} \prod_{t=1}^{T} \int_{-\infty}^{+\infty} \exp\left[-\frac{s_t^2}{2\alpha^2} - \frac{(h_{1t} - s_t)^2}{2\sigma_1^2} - \frac{(h_{2t} - s_t)^2}{2\sigma_2^2}\right] \\ & \times \frac{1}{\sqrt{2\pi}\alpha} g_s(x)\, g_{n_1}(x)\, g_{n_2}(x)\, ds_t \end{aligned} \tag{5.79}$$

$$\begin{aligned} = & \max_{\substack{\alpha, \sigma_1, \sigma_2, c_3, c_4, \\ c_{1,3}, c_{1,4}, c_{2,3}, c_{2,4}}} \frac{1}{(2\pi\sigma_1\sigma_2)^T} \prod_{t=1}^{T} \exp\left[-\frac{h_{1t}^2}{2\sigma_1^2} - \frac{h_{2t}^2}{2\sigma_2^2}\right] \exp\left[\frac{1}{2}\sigma^2 \left(\frac{h_{1t}}{\sigma_1^2} + \frac{h_{2t}}{\sigma_2^2}\right)^2\right] \\ & \times \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(s_t - \mu_t)^2\right] g_s(s_t)\, g_{n_1}(s_t)\, g_{n_2}(s_t)\, ds_t. \end{aligned} \tag{5.80}$$

with:

$$\sigma = \left(\frac{1}{\alpha^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-\frac{1}{2}}, \tag{5.81}$$

$$\mu_t = \left(\frac{h_{1t}}{\sigma_1^2} + \frac{h_{2t}}{\sigma_2^2}\right)\sigma^2. \tag{5.82}$$

Focusing for simplicity on symmetric noise distribution functions, for which we have $c_3, c_{1,3}, c_{2,3} = 0$, we obtain:

$$g_s(x)\, g_{n_1}(x)\, g_{n_2}(x) = \left(b_0 + b_2 x^2 + b_4 x^4\right)\left(b_{1,0} + b_{1,2} x^2 + b_{1,4} x^4\right)\left(b_{2,0} + b_{2,2} x^2 + b_{2,4} x^4\right) \tag{5.83}$$

$$= \gamma_0 + \gamma_2 x^2 + \gamma_4 x^4 + \gamma_6 x^6 + \gamma_8 x^8 + \gamma_8 x^{10} + \gamma_8 x^{12}, \tag{5.84}$$

with straightforward expressions for the γ_i as a function of the $b_{1,i}$, $b_{2,i}$ and b_i coefficients. We therefore need to compute the following integrals, which can be obtained from the first moments of the Gaussian distribution:

$$I_0 = \gamma_0 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}\left(s_t - \mu_t\right)^2\right] ds_t \tag{5.85}$$

$$I_{2t} = \gamma_2 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t^2 \exp\left[-\frac{1}{2\sigma^2}\left(s_t - \mu_t\right)^2\right] ds_t \tag{5.86}$$

$$I_{4t} = \gamma_4 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t^4 \exp\left[-\frac{1}{2\sigma^2}\left(s_t - \mu\right)^2\right] ds_t \tag{5.87}$$

$$I_{6t} = \gamma_6 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t^6 \exp\left[-\frac{1}{2\sigma^2}\left(s_t - \mu_t\right)^2\right] ds_t \tag{5.88}$$

$$I_{8t} = \gamma_8 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t^8 \exp\left[-\frac{1}{2\sigma^2}\left(s_t - \mu_t\right)^2\right] ds_t \tag{5.89}$$

$$I_{10t} = \gamma_{10} \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t^{10} \exp\left[-\frac{1}{2\sigma^2}\left(s_t - \mu_t\right)^2\right] ds_t \tag{5.90}$$

$$I_{12t} = \gamma_{12} \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} s_t^8 \exp\left[-\frac{1}{2\sigma^2}\left(s_t - \mu_t\right)^2\right] ds_t \tag{5.91}$$

Again, we need to compute higher-order moments of the Gaussian distribution, which is given by the following formula for a normally distributed variable X with mean μ_t and variance σ^2:

$$\mathbb{E}\left(X^n\right) = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} x^n \exp\left[-\frac{1}{2\sigma^2}\left(x-\mu_t\right)^2\right] dx = \sum_{j=0}^{\left[\frac{n}{2}\right]} \binom{n}{2j} (2j-1)!!\sigma^{2j}\mu^{n-2j}, \quad (5.92)$$

where $n!!$ denotes the double factorial operator $n!! = \prod_{i=0}^{k}(n-2i) = n\left(n-2\right)\left(n-4\right)...$ with $k = \left[\frac{n}{2}\right]$. For example, we have that:

$$I_{10t} = \gamma_{10}\left(\mu_t^{10} + 45\mu_t^8\sigma^2 + 630\mu_t^6\sigma^4 + 3150\mu_t^4\sigma^6 + 4725\mu_t^2\sigma^8 + 945\sigma^{10}\right). \quad (5.93)$$

Finally, we obtain for the numerator N of the detection statistic Λ_{ML} (again similar expressions can be obtained for the denominator D which relates to the simpler situation when there is no signal):

$$\begin{aligned} N \quad = \quad & \max_{\substack{\alpha,\,\sigma_1,\,\sigma_2,\,c_3,\,c_4,\\ c_{1,3},\,c_{1,4},\,c_{2,3},\,c_{2,4}}} \frac{1}{\left(2\pi\sigma_1\sigma_2\right)^T} \prod_{t=1}^{T} \frac{\sigma}{\alpha} \exp\left[-\frac{h_{1t}^2}{2\sigma_1^2} - \frac{h_{2t}^2}{2\sigma_2^2}\right] \exp\left[\frac{1}{2}\sigma^2\left(\frac{h_{1t}}{\sigma_1^2} + \frac{h_{2t}}{\sigma_2^2}\right)^2\right] \\ & \times \left(I_0 + I_{1t} + I_{2t} + I_{3t} + I_{4t} + I_{6t} + I_{8t} + I_{10t} + I_{12t}\right). \end{aligned} \quad (5.94)$$

We note that when $c_{i,3} = c_{i,4} = 0$, that is when the third and fourth-order cumulants vanish for the noise distribution, we then recover the expression from the previous chapter.

5.2.4 Derivation of the Optimal Cross-Correlation Statistic in the Presence of Non-Gaussian Noise and SGWB Distributions

In what follows, we discuss the implications of the results obtained thus far for the derivation of an optimal detection statistic. Developing a full-fledged numerical analysis of the sensitivity gains involved in the use of this optimal non-Gaussian detection statistic versus the standard cross-correlation statistic is beyond the scope of this chapter, and is left for further research.

The Case of Co-Aligned and Co-Located Detectors

In chapter 3 (see equation 3.38), we recall that the likelihood ratio in the Gaussian case is given by:

$$\Lambda_{ML} = \max_{\alpha,\sigma_1,\sigma_2\geq 0} \left\{ \frac{\overline{\sigma}_1\overline{\sigma}_2}{\sqrt{\sigma_1^2\sigma_2^2+\sigma_1^2\alpha^2+\sigma_2^2\alpha^2}} \exp\left[\frac{\frac{\overline{\sigma}_1^2}{\sigma_1^4}+\frac{\overline{\sigma}_2^2}{\sigma_2^4}+\frac{2\overline{\alpha}^2}{\sigma_1^2\sigma_2^2}}{2\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}+\frac{1}{\alpha^2}\right)} - \frac{\overline{\sigma}_1^2}{2\sigma_1^2} - \frac{\overline{\sigma}_2^2}{2\sigma_2^2} + 1 \right] \right\}^T . \tag{5.95}$$

When plugging the maximum likelihood estimators $\widehat{\alpha}$, $\widehat{\sigma}_1$ and $\widehat{\sigma}_2$ in this equation we obtain the corresponding detection statistic:

$$\Lambda_{ML}^G = \left(\frac{\overline{\sigma}_1\overline{\sigma}_2}{\sqrt{\widehat{\sigma}_1^2\widehat{\sigma}_2^2+\widehat{\sigma}_1^2\widehat{\alpha}^2+\widehat{\sigma}_2^2\widehat{\alpha}^2}} \right)^T \exp T \left[\frac{\frac{\overline{\sigma}_1^2}{\widehat{\sigma}_1^4}+\frac{\overline{\sigma}_2^2}{\widehat{\sigma}_2^4}+\frac{2\overline{\alpha}^2}{\widehat{\sigma}_1^2\widehat{\sigma}_2^2}}{2\left(\frac{1}{\widehat{\sigma}_1^2}+\frac{1}{\widehat{\sigma}_2^2}+\frac{1}{\widehat{\alpha}^2}\right)} - \frac{\overline{\sigma}_1^2}{2\widehat{\sigma}_1^2} - \frac{\overline{\sigma}_2^2}{2\widehat{\sigma}_2^2} + 1 \right], \tag{5.96}$$

which simplifies into:

$$\Lambda_{ML}^G = \left(1 - \frac{\widehat{\alpha}^4}{\overline{\sigma}_1^2\overline{\sigma}_2^2} \right)^{-T/2} . \tag{5.97}$$

This expression can be regarded as the optimal detection statistic by Neyman-Pearson theorem. In practice, we instead use the following transformation, known as the cross-

correlation detection statistic:

$$DS_{cc} = \sqrt{T}\frac{\overline{\alpha}^2}{\overline{\sigma}_1\overline{\sigma}_2} \tag{5.98}$$

This methodology can be extended to account for the presence of deviations from the Gaussian assumption for both the signal and noise distributions since we have $\Lambda_{ML} = \frac{N}{D}$, where:

$$\begin{aligned} N \;=\; & \max_{\substack{\alpha,\sigma_1,\sigma_2,c_3,c_4,\\ c_{1,3},c_{1,4},c_{2,3},c_{2,4}}} \frac{1}{(2\pi\sigma_1\sigma_2)^T}\prod_{t=1}^{T}\frac{\sigma}{\alpha}\exp\left[-\frac{h_{1t}^2}{2\sigma_1^2}-\frac{h_{2t}^2}{2\sigma_2^2}\right]\exp\left[\frac{1}{2}\sigma^2\left(\frac{h_{1t}}{\sigma_1^2}+\frac{h_{2t}}{\sigma_2^2}\right)^2\right] \\ & \times\left(I_0+I_{1t}+I_{2t}+I_{3t}+I_{4t}+I_{6t}+I_{8t}+I_{10t}+I_{12t}\right). \end{aligned} \tag{5.99}$$

Let us denote $I_t \equiv I_0 + I_{1t} + I_{2t} + I_{3t} + I_{4t} + I_{6t} + I_{8t} + I_{10t} + I_{12t}$. Through the $\gamma_1, ..., \gamma_{10}$ coefficients I_t is a function of the parameters $c_3, c_4, c_{1,3}, c_{1,4}, c_{2,3}, c_{2,4}$:

$$I_t = f\left(c_3, c_4, c_{1,3}, c_{1,4}, c_{2,3}, c_{2,4}\right).$$

We note again that when $c_3 = c_4 = c_{1,3} = c_{1,4} = c_{2,3} = c_{2,4} = 0$, that is when the third and fourth-order cumulants vanish for the signal and noise distributions, we then recover the standard maximum likelihood statistic. In the general case however the optimal non-Gaussian detection statistic is given by:

$$\Lambda_{ML}^{NG} = \frac{N\left(\widehat{\alpha},\widehat{c}_3,\widehat{c}_4,\widehat{\sigma}_1,\widehat{\sigma}_2,\widehat{c}_{1,3},\widehat{c}_{1,4},\widehat{c}_{2,3},\widehat{c}_{2,4}\right)}{D\left(\overline{\sigma}_1,\overline{\sigma}_2,\overline{c}_{1,3},\overline{c}_{1,4},\overline{c}_{2,3},\overline{c}_{2,4}\right)} \tag{5.100}$$

where $\widehat{\alpha},\widehat{c}_3,\widehat{c}_4,\widehat{\sigma}_1,\widehat{\sigma}_2,\widehat{c}_{1,3},\widehat{c}_{1,4},\widehat{c}_{2,3},\widehat{c}_{2,4}$ are the values of the corresponding parameters that maximize the numerator of the likelihood function, while $\overline{\sigma}_1,\overline{\sigma}_2,\overline{c}_{1,3},\overline{c}_{1,4},\overline{c}_{2,3},\overline{c}_{2,4}$ are the values of the corresponding parameters that maximize the denominator of the likelihood function. It is important to emphasize that in this general non-Gaussian setting, the optimal detection statistic given in 5.100 is a priori different from the signal-to-noise ratio.

A key difference with the Gaussian case, however, is that these estimators are not available in analytical form in the general non-Gaussian setting. From the analysis presented in chapter 4, it turns out that we may in fact derive analytical expressions for competing estimators called moment-based estimators. The benefit of using moment-based estimators is that the optimal detection statistic can then be obtained in closed-form and numerical estimates for the sensitivity gain can be performed for realistic parameter values. To derive these moment-based estimators, we write (focusing here on even moments of the noise and signal distributions):

$$\mathbb{E}\left(\mathcal{H}_1\mathcal{H}_2\right) = \mathbb{E}\left[\left(\mathcal{N}_1+\mathcal{S}\right)\left(\mathcal{N}_2+\mathcal{S}\right)\right] = \mathbb{E}\left[\mathcal{S}^2\right] = \alpha^2 \tag{5.101}$$

$$\mathbb{E}\left(\mathcal{H}_1^2\mathcal{H}_2^2\right) = \mathbb{E}\left[\left(\mathcal{N}_1+\mathcal{S}\right)^2\left(\mathcal{N}_2+\mathcal{S}\right)^2\right] = c_4 + 3\alpha^2 + \alpha^2\left(\sigma_1^2+\sigma_2^2\right) + \sigma_1^2\sigma_2^2 \tag{5.102}$$

$$\mathbb{E}\left(\mathcal{H}_1^4\right) = \mathbb{E}\left[\left(\mathcal{N}_1+\mathcal{S}\right)^4\right] = c_4 + 3\alpha^2 + 6\alpha^2\sigma_1^2 + c_{1,4} + 3\sigma_1^2 \tag{5.103}$$

$$\mathbb{E}\left(\mathcal{H}_2^4\right) = \mathbb{E}\left[\left(\mathcal{N}_2+\mathcal{S}\right)^4\right] = c_4 + 3\alpha^2 + 6\alpha^2\sigma_2^2 + c_{2,4} + 3\sigma_2^2 \tag{5.104}$$

From this analysis, we obtain the corresponding moment-based estimators to be used in the numerator N of the detection statistic:

$$\widehat{\alpha}^2 = \frac{1}{T}\sum_{t=1}^{T}\mathcal{H}_{1t}\mathcal{H}_{2t} \tag{5.105}$$

$$\widehat{\sigma}_1^2 = \frac{1}{T}\sum_{t=1}^{T}\mathcal{H}_{1t}^2 - \widehat{\alpha}^2 \tag{5.106}$$

$$\widehat{\sigma}_2^2 = \frac{1}{T}\sum_{t=1}^{T}\mathcal{H}_{2t}^2 - \widehat{\alpha}^2 \tag{5.107}$$

$$\widehat{c}_4 = \frac{1}{T}\sum_{t=1}^{T}\mathcal{H}_{1t}^2\mathcal{H}_{2t}^2 - 3\widehat{\alpha}^2 - \widehat{\alpha}^2\left(\widehat{\sigma}_1^2+\widehat{\sigma}_2^2\right) - \widehat{\sigma}_1^2\widehat{\sigma}_2^2 \tag{5.108}$$

$$\widehat{c}_{1,4} = \frac{1}{T}\sum_{t=1}^{T}\mathcal{H}_{1t}^4 - \widehat{c}_4 - 3\widehat{\alpha}^2 - 6\widehat{\alpha}^2\widehat{\sigma}_1^2 - 3\widehat{\sigma}_1^2 \tag{5.109}$$

$$\widehat{c}_{2,4} = \frac{1}{T}\sum_{t=1}^{T}\mathcal{H}_{2t}^4 - \widehat{c}_4 - 3\widehat{\alpha}^2 - 6\widehat{\alpha}^2\widehat{\sigma}_2^2 - 3\widehat{\sigma}_2^2 \tag{5.110}$$

To obtain the corresponding expressions $\overline{\sigma}_1, \overline{\sigma}_2, \overline{c}_{1,4}, \overline{c}_{2,4}$ to be used in the denominator

D of the detection statistic, one simply takes $\widehat{\alpha}^2 = 0$ and $\widehat{c}_4 = 0$ in the equations 5.105, 5.106, 5.109, and 5.110 giving the estimators for σ_1^2, σ_2^2, $c_{1,4}$ and $c_{2,4}$.

In the Gaussian case, the moment-based estimators in equations 5.105, 5.106 and 5.107 for the variance of the signal and the noise in each detector coincide with the maximum likelihood estimators. In the general non-Gaussian case the moment-based estimators may not coincide with the maximum likelihood estimators but at least they enjoy the key desirable property to be unbiased by construction. The analysis conducted in chapter 4 (see Table 4.2 and Figure 4.7) for a situation with a non-Gaussian signal and Gaussian noise actually suggests that the moment-based estimators are only marginally less efficient as the maximum-likelihood estimators (meaning they have a slightly higher variance), which justifies their use in the derivation of the optimal detection statistic. Plugging these expressions back in equation 5.100, we can thus obtain the expression for the optimal detection statistic in the non-Gaussian case as a function of the joint observations $h = (h_{1t}, h_{2t})_{0 \leq t \leq T}$. We note that when the signal and noise distributions are Gaussian, that is when $\widehat{c}_4, \widehat{c}_{1,4}, \widehat{c}_{2,4} = 0$ then $I_t = 1$ for all t and the optimal detection statistic Λ_{ML}^{NG} coincides with the standard Gaussian detection statistic that it nests as a special case. Based on realistic parameter values, one can in principle numerically measure the sensitivity gain involved in using in the non-Gaussian case the optimal non-Gaussian statistic as opposed to using the standard cross-correlation statistic, which is only optimal when signal and noise distributions are Gaussian.

Extending the Results to the Case of Separated Detectors

As indicated before, the cross-correlation detection statistic may not be optimal in the general non-Gaussian case, and one would like to derive an optimal detection statistic when the higher order cumulants are not zero for possibly both the signal and noise distributions. As recalled in chapter 3, the optimal choice of a filter function depends on the optimization objective. In what follows, we extend the analysis to the case of separated detectors through a filtering approach that focuses on maximizing the signal

to noise ratio:

$$\frac{S}{N} = \frac{\mathbb{E}\left(Y\right)}{\sqrt{\mathbb{V}ar\left(Y\right)}}. \tag{5.111}$$

Under aforementioned standard assumptions, one can show [17] that the optimization criterion defined by minimization of the probability of a false alarm (pfa) subject to a fixed probability of a false dismissal (pfd) is equivalent to maximizing the signal to noise ratio of the cross-correlation detection statistic. In a general context, however, the maximization of a signal to noise ratio involving the first two moments of the cross-correlation statistic will not be a priori equivalent to minimizing the probability of a false alarm for a given probability of a false dismissal. For tractability, we nonetheless maintain here the focus on maximizing the signal to noise ratio, and we show in what follows that the use of the Edgeworth expansion allows us to maintain an analytical expression for the optimal filter function in this case. More precisely, we first write the general expression for the cross-correlation product Y for an isotropic, unpolarized and stationary SGWB in the case of co-aligned and co-located detectors, while we discuss the situation with separated detectors below. We have:

$$Y = \int_{-\infty}^{+\infty} \widetilde{s}_1^{*}\left(f\right) \widetilde{Q}\left(f\right) \widetilde{s}_2\left(f\right) df, \tag{5.112}$$

where the optimal filtering function $\widetilde{Q}$ is defined as the quantity that maximizes the signal to noise ratio. One can easily check that the derivation of this quantity in 3.73 does not depend upon the assumption that the signal or noise distributions were Gaussian, and the following standard expression for the optimal filter remains valid:

$$\widetilde{Q}\left(f\right) = \frac{\lambda \Omega_{GW}\left(f\right)}{f^3 S_{n_1}\left(f\right) S_{n_2}\left(f\right)}, \tag{5.113}$$

with $S_{n_1}\left(f\right)$ and $S_{n_2}\left(f\right)$ being the power spectral noise densities of the two detectors,

where λ is a normalization constant and where (equation 3.53):

$$\Omega_{GW}(f) = \frac{4\pi^2}{3H_0^2} f^3 S_s(f). \tag{5.114}$$

We finally have:

$$\widetilde{Q}(f) = \frac{4\pi^2}{3H_0^2} \frac{S_s(f)}{S_{n_1}(f) S_{n_2}(f)} \tag{5.115}$$

The only difference between the Gaussian and non-Gaussian case relates to a different expression for the Fourier transform of the spectral density $S_s(f)$ of the signal distribution and the spectral density $S_{n_1}(f)$ and $S_{n_2}(f)$ of the noise distributions. These expressions are in turn related to the Fourier transforms of the signal and noise distributions, and we show below that these Fourier transforms can be obtained in closed-form when the unknown signal and noise distribution functions are approximated by means on the Edgeworth expansion. Indeed, the Fourier transform of the approximate density can be obtained as the Fourier transform of the product $f_s(x) = \phi_s(x) g_s(x)$ and $f_{n_i}(x) = \phi_{n_i}(x) g_{n_i}(x)$ for the signal and noise distributions, where $g_s(x)$ and $g_{n_i}(x)$ are polynomial functions. Focusing for example on the polynomial correction to the signal probability density, we have in the case of a symmetric distribution ($c_3 = 0$):

$$\begin{aligned} g_s(x) &= 1 + \frac{c_4}{8\alpha^4} - \frac{c_4}{4\alpha^6} x^2 + \frac{c_4}{24\alpha^8} x^4 && (5.116) \\ &\equiv a_{0,s} + a_{2,s} x^2 + a_{4,s} x^4. && (5.117) \end{aligned}$$

Introducing the elementary polynomial $P_n(x) = x^n$, we obtain $f_s(x) \simeq \sum_{n=0,2,4} a_{n,s} P_n(x) \phi_s(x)$. Since the Fourier transform of the sum of functions $a_n P_n(x) \phi_s(x)$ is the sum of the Fourier transforms of the functions $a_n P_n(x) \phi_s(x)$, we are therefore left with the computation of the Fourier transform of the product $\phi_s(x) P_n(x)$. This Fourier transform is

well-known to be given by the *convolution* of the Fourier transforms of $\phi_s(x)$ and $P_n(x)$:

$$\widetilde{(\phi_s \times P_n)}(f) = \widetilde{\phi}_s(f) * \widetilde{P_n}(f) \equiv \int_{-\infty}^{+\infty} \widetilde{\phi}_s(q)\, \widetilde{P_n}(f-q)\, dq. \tag{5.118}$$

We therefore have:

$$\widetilde{f}_s(f) = \sum_{n=0,2,4} a_{n,s} \widetilde{(\phi_s \times P_n)}(f) = \sum_{n=0,2,4} a_{n,s} \left[\widetilde{\phi}_s(f) * \widetilde{P_n}(f)\right]. \tag{5.119}$$

It is well-known that the Fourier transform of a Gaussian is also a Gaussian and given by:

$$\phi_s(x) = \frac{1}{\sqrt{2\pi}\alpha} e^{-\frac{x^2}{2\alpha^2}} \Rightarrow \widetilde{\phi}_s(f) = \int_{-\infty}^{+\infty} e^{-2\pi i f t} \phi_s(x)\, dx = \frac{\alpha}{\sqrt{2\pi}} e^{-\frac{f^2\alpha^2}{2}}. \tag{5.120}$$

We also have:

$$P_n(x) = x^n \Rightarrow \widetilde{P_n}(f) = \int_{-\infty}^{+\infty} e^{-2\pi i f t} P_n(x)\, dx = i^n \sqrt{2\pi}\delta^{(n)}(f), \tag{5.121}$$

where $\delta^{(n)}$ is the n^{th} distribution derivative of the Dirac delta function. From fundamental properties of the Fourier transform and of the Dirac distributions, we know that

$$(\delta') * f = -\delta * f' = -f'. \tag{5.122}$$

Similarly,

$$\delta^{(n)} * f = (-1)^n \delta * f^{(n)} = (-1)^n f^{(n)}, \tag{5.123}$$

so we finally obtain:

$$\widetilde{(\phi_s \times P_n)}(f) = (-1)^n \alpha^{2n} \widetilde{\phi}_s(f). \tag{5.124}$$

This relationship can be extended to the polynomial $g_s(x) = \sum a_n x^n$ so as to give

$$\widetilde{f}_s(f) = \widetilde{(\phi_s \times g_s)}(f) = \sum_{n=0,2,4} (-1)^n a_{n,s} \alpha^{2n} \widetilde{\phi}_s(f). \tag{5.125}$$

From these results, we can obtain analytical expression for the signal spectral density $S_s(f) = \widetilde{f}_s^{\,*}(f)\,\widetilde{f}_s(f)$, which is one of the required inputs in the expression for the optimal filter 5.115. Similar expressions can be obtained for the Fourier transforms of the noise density functions $f_{n_i}(x)$, and therefore for the corresponding noise spectral density functions $S_{n_1}(f)$ and $S_{n_2}(f)$.

We have considered so far colocated and coincident detectors, an assumption which would hold in the case of Einstein Telescope. On the other hand, our framework should be extended to apply to a network of separated detectors such as in the early phases of development of Advanced LIGO-Virgo detectors, or joint observations by Advanced LIGO and Einstein Telescope. This extension is important because these are precisely the types of situations where differences in sensitivities are expected to be substantial. As recalled in chapter 3, when the two detectors are not colocated and/or not coaligned, we expect a reduction in sensitivity implying that the expression for $\widetilde{Q}(f)$ is transformed as follows (equation 3.73):

$$\widetilde{Q}(f) = \frac{\gamma(|f|)\,\Omega_{GW}(|f|)}{|f|^3\, S_{n_1}(|f|)\, S_{n_2}(|f|)}, \tag{5.126}$$

where $\gamma(f)$ is the (normalized) *overlap reduction function*. Because the Edgeworth expansion allows one to write the approximate density of the non-Gaussian distribution for the signal and/or detector noise as the product of a polynomial of degree 4 (case of a symmetric distribution) or 6 (case of a non-symmetric distribution), one can again easily extend the standard analysis performed in the Gaussian situation by taking the Fourier transform of the approximate non-Gaussian densities as explained above, obtain analytical expression for the signal and noise spectral densities, and use them as inputs in 5.126 for the extended optimal detection statistic for non-Gaussian signal and/or noise

distributions with detectors that are not colocated and coaligned.

6

Conclusions and Perspectives

This thesis proposes a formal analysis of detection and estimation methods for gravitational wave stochastic backgrounds in non-Gaussian regimes. In a first step, we use the Edgeworth expansion, which is a formal expansion of the characteristic function of the signal distribution whose unknown probability density function is to be approximated in terms of the characteristic function of the Gaussian distribution. The non-Gaussian estimation procedure we obtain generalizes the standard cross-correlation statistic, which is recovered in the limit of vanishing third and fourth cumulants of the empirical conditional distribution of the detector measurement. Our research complements related work [62], where the authors have focused on a very specific model where the deviation from the Gaussian distributional assumption was understood as emanating from the presence of a resolved Gaussian signal being measured. We provide a methodology that can be applied without any assumption regarding the exact origin of the departure from normality, and which relies on an explicit semi-parametric approximation of the unknown density function. The main benefit of the procedure is that it allows us to estimate additional parameters when the signal is not too small compared to the noise (signal to noise ratio of the order of 1%), namely the 3rd and 4th cumulant of the gravitational wave signal distribution. In a second step, we analyze the efficiency of the cross-correlation

statistic in situations that deviate from the Gaussian assumption for both the stochastic gravitational wave signal distribution and detector noise distributions. To do so we show how to obtain consistent estimates for the first four cumulants of the signal and noise distributions using a suitable extension of the likelihood function, for which we derive an analytical expression. In a numerical analysis, we find that properly accounting for the presence of non-Gaussian distributions as opposed to wrongly assuming that higher-order cumulants of the noise distributions are zero has material implications in the implementation of standard detection procedures in that it generates higher values for probabilities of false dismissal corresponding to given levels of probabilities of false alarm. The required correction is found to be particularly large when detector sensitivities exhibit substantial differences, a situation that is expected to hold in early phases of development of the Advanced LIGO-Virgo detectors before they reach their design sensitivity or in joint detections from Advanced LIGO and the Einstein Telescope project [142]. These results are obtained in a specialized setting where detector sensitivities are assumed to be known or at least sufficiently well-estimated, and they would not be valid in realistic situations when the uncertainty in the estimated values for the detector noise is larger than the size of the stochastic signal to be detected. In addition to outlining their implications for the performance of the standard CC detection statistic, we also discuss the implications of these results for the derivation of an optimal detection statistic in a non-Gaussian context.

In summary, our work has potential practical implications for both the detection of SGWB signals and the estimation of related parameters. In terms of parameter estimation, our methods based on generalized expressions for standard maximum likelihood estimators can prove useful because they precisely allow for the efficient estimation of additional parameters, namely the higher-order cumulants of the signal and/or noise distributions, which should have a number of useful astrophysical implications. Generating estimates for additional parameters should allow us in particular to obtain additional constraints on astrophysical and cosmological models that will be imposed by observed

gravitation wave signals, and comparing them to the constraints derived from supernovae or galaxy clusters observations. For instance, using an isotropic search conducted on mock data for compact binary mergers as a function of the simulated average strength of the signal over the observational period, [126] find that a high coalescence rate of events and low average chirp mass or a low rate of events and high average mass generate the same signal amplitude and spectral index for the frequency range under analysis. Our non-Gaussian estimation methodologies can in principle be used to help differentiate these two types of signals, a continuous signal resulting in a low kurtosis in the first high rate/low mass case, and a popcorn-like signal resulting in a high kurtosis in the second low rate/high mass situations. Turning to the detection problem, we introduce a framework for deriving an optimal detection statistic in a generalized non-Gaussian setting, which nests the standard CC detection statistic as a specific case when the third and fourth cumulants of the signal and noise distributions are zero. Developing a full-fledged numerical analysis of the sensitivity gains involved in the use of the optimal non-Gaussian detection statistic versus the standard cross-correlation statistic would be a natural extension of the work presented in this thesis.

The results presented here can be extended in a number of additional useful directions. One possible extension of our work would consist in trying to provide a characterization of the non-Gaussian signature of the SGWB that could help distinguish between its astrophysical versus cosmological GW origin. For example, one may try and obtain a theoretical estimation for the cumulant parameters of a distribution sourced by cosmic string events with a focus on LIGO-Virgo frequencies. As recalled above, Damour and Vilenkin [58] made the observation that the stochastic ensemble of GWs generated by a network of cosmic strings includes large infrequent bursts, and that the computation of $\Omega_{GW}(f)$ should not be biased by including these large rare events. When loops are small, all loops at a certain redshift are the same size and produce the same amplitude events. Hence, a cutoff can be placed in the integral over redshifts to remove large events for which the rate is smaller than the relevant time-scale of the experiment (see equation

(6.17) of [58]). When loops are large, there can be loops of any size at any given redshift. One can deal with this problem by evaluating the rate from cusps in redshift interval dz and with strain in the interval dh [160] to remove large amplitude events, defined as events characterized by a strain amplitude that exceeds a given threshold which is a function of the detector sensitivity, that occur at a rate smaller than f. This procedure has been used in [160] to find the areas of parameter space of cosmic string models that make a detection possible with Advanced LIGO, LISA and pulsar timing experiments. In principle, we could use a similar approach to analyze the Gaussian or non-Gaussian nature of the SGWB signal distribution. As we let the finite detector sensitivity increase within reasonable bounds, the number of resolved detected GW signals is expected to increase, and the distribution of the residual background is expected to become more non-Gaussian, as can be characterized with a numerical estimate for the 3rd and 4th moment of its distribution. We could then potentially contrast this estimation with an estimation of cumulants that would be obtained from realistic deviations from the Central Limit Theorem for astrophysical sources of GW stochastic background. If the cosmological and astrophysical backgrounds were to produce somewhat different types of non-normality, this would lead to empirically testable predictions that could be validated using our approach.

Bibliography

[1] J Aasi, J Abadie, BP Abbott, R Abbott, T Abbott, MR Abernathy, T Accadia, F Acernese, C Adams, T Adams, et al. Gravitational waves from known pulsars: results from the initial detector era. *The Astrophysical Journal*, 785(2):119, 2014.

[2] J Aasi, J Abadie, BP Abbott, R Abbott, TD Abbott, M Abernathy, T Accadia, F Acernese, C Adams, T Adams, et al. Prospects for localization of gravitational wave transients by the advanced ligo and advanced virgo observatories. *Living Reviews in Relativity*, 19, 2016.

[3] Junaid Aasi, BP Abbott, Richard Abbott, Thomas Abbott, MR Abernathy, Kendall Ackley, Carl Adams, Thomas Adams, Paolo Addesso, RX Adhikari, et al. Advanced ligo. *Classical and Quantum Gravity*, 32(7):074001, 2015.

[4] BP Abbott, R Abbott, TD Abbott, MR Abernathy, F Acernese, K Ackley, C Adams, T Adams, P Addesso, RX Adhikari, et al. Binary black hole mergers in the first advanced ligo observing run. *arXiv preprint arXiv:1606.04856*, 2016.

[5] BP Abbott, R Abbott, TD Abbott, MR Abernathy, F Acernese, K Ackley, C Adams, T Adams, P Addesso, RX Adhikari, et al. Gw151226: Observation of gravitational waves from a 22-solar-mass binary black hole coalescence. *Phys. Rev. Lett.*, 116:241103, Jun 2016.

[6] BP Abbott, R Abbott, TD Abbott, MR Abernathy, F Acernese, K Ackley, C Adams, T Adams, P Addesso, RX Adhikari, et al. Observing gravitational-wave transient gw150914 with minimal assumptions. *Physical Review D*, 93(12):122004, 2016.

[7] BP Abbott, R Abbott, R Adhikari, P Ajith, Bruce Allen, G Allen, RS Amin, SB Anderson, WG Anderson, MA Arain, et al. Ligo: the laser interferometer gravitational-wave observatory. *Reports on Progress in Physics*, 72(7):076901, 2009.

[8] BP Abbott, Richard Abbott, TD Abbott, MR Abernathy, Fausto Acernese, Kendall Ackley, Carl Adams, Thomas Adams, Paolo Addesso, RX Adhikari, et al. Gw150914: Implications for the stochastic gravitational-wave background from binary black holes. *Physical review letters*, 116(13):131102, 2016.

[9] BP Abbott, Richard Abbott, TD Abbott, MR Abernathy, Fausto Acernese, Kendall Ackley, Carl Adams, Thomas Adams, Paolo Addesso, RX Adhikari, et al. Observation of gravitational waves from a binary black hole merger. *Physical review letters*, 116(6):061102, 2016.

[10] Rana X Adhikari. Gravitational radiation detection with laser interferometry. *Reviews of Modern Physics*, 86(1):121, 2014.

[11] Odylio Denys Aguiar. Past, present and future of the resonant-mass gravitational wave detectors. *Research in Astronomy and Astrophysics*, 11(1):1, 2011.

[12] Hassan Akbar-Zadeh. *Initiation to global Finslerian geometry*, volume 68. Elsevier, 2006.

[13] B. Allen, J. D. E. Creighton, É. É. Flanagan, and J. D. Romano. Robust statistics for deterministic and stochastic gravitational waves in non-Gaussian noise: Frequentist analyses. *Physical Review D*, 65(12):122002, June 2002.

[14] B. Allen, J. D. E. Creighton, É. É. Flanagan, and J. D. Romano. Robust statistics for deterministic and stochastic gravitational waves in non-Gaussian noise. II. Bayesian analyses. *Physical Review D*, 67(12):122002, June 2003.

[15] B. Allen and J. D. Romano. Detecting a stochastic background of gravitational radiation: Signal processing strategies and sensitivities. *Physical Review D*, 59(10):102001, May 1999.

[16] Bruce Allen, Jolien DE Creighton, Éanna É Flanagan, and Joseph D Romano. Robust statistics for deterministic and stochastic gravitational waves in non-gaussian noise: Frequentist analyses. *Physical Review D*, 65(12):122002, 2002.

[17] Bruce Allen and Joseph D Romano. Detecting a stochastic background of gravitational radiation: Signal processing strategies and sensitivities. *Physical Review D*, 59(10):102001, 1999.

[18] ZA Allen, P Astone, L Baggio, D Busby, M Bassan, DG Blair, M Bonaldi, P Bonifazi, P Carelli, M Cerdonio, et al. First search for gravitational wave bursts with a network of detectors. *Physical Review Letters*, 85(24):5046, 2000.

[19] Luca Amendola. The dependence of cosmological parameters estimated from the microwave background on non-gaussianity. *The Astrophysical Journal*, 569(2):595, 2002.

[20] Theodore W Anderson and Donald A Darling. Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics*, pages 193–212, 1952.

[21] Warren G Anderson and Jolien DE Creighton. Searches for gravitational waves from binary neutron stars: A review. In *Short-Period Binary Stars: Observations, Analyses, and Results*, pages 23–52. Springer, 2008.

[22] M Armano, H Audley, G Auger, JT Baird, M Bassan, P Binetruy, M Born, D Bortoluzzi, N Brandt, M Caleno, et al. Sub-femto-g free fall for space-based gravitational wave observatories: Lisa pathfinder results. *Physical Review Letters*, 116(23):231101, 2016.

[23] PP Avelino, EPS Shellard, JHP Wu, and Bruce Allen. Non-gaussian features of linear cosmic string models. *The Astrophysical Journal Letters*, 507(2):L101, 1998.

[24] B Barr, A Bell, C Bell, C Bond, D Brown, F Brueckner, L Carbone, K Craig, A Cumming, S Danilishin, et al. Ligo 3 strawman design, team red. 2012.

[25] SJ Bell and A Hewish. Angular size and flux density of the small source in the crab nebula at 81.5 mc/s. 1967.

[26] V Berezinsky, B Hnatyk, and A Vilenkin. Gamma ray bursts from superconducting cosmic strings. *Physical Review D*, 64(4):043004, 2001.

[27] Pierre Binétruy, Alejandro Bohé, Chiara Caprini, and Jean-François Dufaux. Cosmological backgrounds of gravitational waves and elisa/ngo: phase transitions, cosmic strings and other sources. *Journal of Cosmology and Astroparticle Physics*, 2012(06):027, 2012.

[28] Rahul Biswas, Patrick R Brady, Jordi Burguet-Castell, Kipp Cannon, Jessica Clayton, Alexander Dietz, Nickolas Fotopoulos, Lisa M Goggin, Drew Keppel, Chris Pankow, et al. Likelihood-ratio ranking of gravitational-wave candidates in a non-gaussian background. *Physical Review D*, 85(12):122008, 2012.

[29] David G Blair. *Advanced gravitational wave detectors*. Cambridge University Press, 2012.

[30] Sergei Blinnikov and Richhild Moessner. Expansions for nearly gaussian distributions. *Astronomy and Astrophysics Supplement Series*, 130(1):193–205, 1998.

[31] George R Blumenthal, SM Faber, Joel R Primack, and Martin J Rees. Formation of galaxies and large scale structure with cold dark matterŠ. 1984.

[32] GE Box, WG Hunter, and JS Hunter. Statistics for engineers: An introduction to design, data analysis, and model building, 1978.

[33] Latham A Boyle and Paul J Steinhardt. Probing the early universe with inflationary gravitational waves. *Physical Review D*, 77(6):063504, 2008.

[34] Alessandra Buonanno. Gravitational waves from the early universe. In *TASI 2002*, pages 855–892. World Scientific, 2004.

[35] Robert R Caldwell, Marc Kamionkowski, and Leven Wadley. The first space-based gravitational-wave detectors. *Physical Review D*, 59(2):027101, 1998.

[36] C. Caprini, R. Durrer, T. Konstandin, and G. Servant. General properties of the gravitational wave spectrum from phase transitions. *Physical Review D*, 79(8):083519, April 2009.

[37] C. Caprini, R. Durrer, and G. Servant. Gravitational wave generation from bubble collisions in first-order phase transitions: An analytic approach. *Physical Review D*, 77(12):124015, June 2008.

[38] C. Caprini, R. Durrer, and G. Servant. The stochastic gravitational wave background from turbulence and magnetic fields generated by a first-order phase transition. *Journal of Cosmology and Astroparticle Physics*, 12:24, December 2009.

[39] Sean M Carroll. *Spacetime and geometry. An introduction to general relativity*, volume 1. 2004.

[40] NASA Goddard Space Flight Center. "the gravitational wave spectrum".

[41] Pisin Chen, Po-Shen Hsin, and Yuezhen Niu. Inflation as a solution to the early universe entropy problem. *arXiv preprint arXiv:1212.1087*, 2012.

[42] Herman Chernoff, EL Lehmann, et al. The use of maximum likelihood estimates in χ^2 tests for goodness of fit. *The Annals of Mathematical Statistics*, 25(3):579–586, 1954.

[43] Adrian Cho. Remembering joseph weber, the controversial pioneer of gravitational waves. *Science Online Edition*, 2016.

[44] N. Christensen. Measuring the stochastic gravitational-radiation background with laser-interferometric antennas. *Physical Review D*, 46:5250–5266, December 1992.

[45] LIGO Scientific Collaboration, Virgo Collaboration, et al. Binary black hole mergers in the first advanced ligo observing run. *arXiv preprint arXiv:1606.04856*, 2016.

[46] Planck Collaboration et al. Planck 2015 results. xiii. cosmological parameters. arxiv preprint. *arXiv*, 1502, 2015.

[47] Edmund A Cornish and Ronald A Fisher. Moments and cumulants in the specification of distributions. *Revue de l'Institut international de Statistique*, pages 307–320, 1938.

[48] Neil J Cornish and Joseph D Romano. Towards a unified treatment of gravitational-wave data analysis. *Physical Review D*, 87(12):122003, 2013.

[49] Neil J Cornish and Joseph D Romano. When is a gravitational-wave signal stochastic? *Physical Review D*, 92(4):042001, 2015.

[50] D. M. Coward and R. R. Burman. A cosmological 'probability event horizon' and its observational implications. *Monthly Notices of the Royal Astronomical Society*, 361:362–368, July 2005.

[51] David M Coward, Ronald R Burman, and David G Blair. Simulating a stochastic background of gravitational waves from neutron star formation at cosmological distances. *Monthly Notices of the Royal Astronomical Society*, 329(2):411–416, 2002.

[52] Harald Cramér. On the composition of elementary errors: First paper: Mathematical deductions. *Scandinavian Actuarial Journal*, 1928(1):13–74, 1928.

[53] Jolien DE Creighton and Warren G Anderson. *Gravitational-wave physics and astronomy: an introduction to theory, experiment and data analysis*. John Wiley & Sons, 2012.

[54] Ralph B D'Agostino. Transformation to normality of the null distribution of g1. *Biometrika*, pages 679–681, 1970.

[55] T. Damour and A. Vilenkin. Gravitational radiation from cosmic (super)strings: Bursts, stochastic background, and observational windows. *Physical Review D*, 71(6):063510, March 2005.

[56] Thibault Damour and Alessandro Nagar. Improved analytical description of inspiralling and coalescing black-hole binaries. *Physical Review D*, 79(8):081503, 2009.

[57] Thibault Damour and Alexander Vilenkin. Gravitational wave bursts from cosmic strings. *Physical Review Letters*, 85(18):3761, 2000.

[58] Thibault Damour and Alexander Vilenkin. Gravitational radiation from cosmic (super) strings: Bursts, stochastic background, and observational windows. *Physical Review D*, 71(6):063510, 2005.

[59] Matthew R DePies and Craig J Hogan. Stochastic gravitational wave background from light cosmic strings. *Physical Review D*, 75(12):125006, 2007.

[60] Robert H Dicke and P James E Peebles. The big bang cosmology-enigmas and nostrums. In *General relativity*. 1979.

[61] Harald Dimmelmeier, Christian D Ott, Andreas Marek, and H-Thomas Janka. Gravitational wave burst signal from core collapse of rotating stars. *Physical Review D*, 78(6):064056, 2008.

[62] S. Drasco and É. É. Flanagan. Detection methods for non-Gaussian gravitational wave stochastic backgrounds. *Physical Review D*, 67(8):082003, April 2003.

[63] RWP Drever. Some new concepts for laser interferometer gravitational wave detectors. In *Dark Matter in Cosmology Quantam Measurements Experimental Gravitation*, volume 1, page 375, 1996.

[64] RWP Drever, John L Hall, FV Kowalski, J Hough, GM Ford, AJ Munley, and H Ward. Laser phase and frequency stabilization using an optical resonator. *Applied Physics B*, 31(2):97–105, 1983.

[65] Freeman J Dyson. Ground-state energy of a finite system of charged particles. *Journal of Mathematical Physics*, 8(8):1538–1545, 1967.

[66] A Einstein. 'the field equations of gravitation'. *On a Heuristic Point of View about the Creation and Conversion of Light 1 On the Electrodynamics of Moving Bodies 10 The Development of Our Views on the Composition and Essence of Radiation 11 The Field Equations of Gravitation 19 The Foundation of the Generalised Theory of Relativity 22*, page 19, 1915.

[67] A Einstein. Zur allgemeinen relativitätstheorie (nachtrag), 11 nov 1915. 1915.

[68] Albert Einstein. Ueber das relativitätsprinzip und die aus demselben gezogenen folgerungen,] ahrbuch der radioaktivität, iv, 1907, 411-462; v, 1908, 98-99 (berichtigungen, errata). republ.

[69] Albert Einstein. Ist die trägheit eines körpers von seinem energieinhalt abhängig? *Annalen der Physik*, 323(13):639–641, 1905.

[70] Albert Einstein. Zur elektrodynamik bewegter körper. *Annalen der physik*, 322(10):891–921, 1905.

[71] Albert Einstein. Erklarung der perihelionbewegung der merkur aus der allgemeinen relativitatstheorie. *Sitzungsber. preuss. Akad. Wiss., vol. 47, No. 2, pp. 831-839, 1915*, 47:831–839, 1915.

[72] Albert Einstein. Grundgedanken der allgemeinen relativitätstheorie und anwendung dieser theorie in der astronomie. *Preussische Akademie der Wissenschaften, Sitzungsberichte*, 315, 1915.

[73] Albert Einstein. *Näherungsweise integration der feldgleichungen der gravitation.* Wiley Online Library, 1916.

[74] Albert Einstein. Kosmologische betrachtungen zur allgemeinen relativitätstheorie. In *Das Relativitätsprinzip*, pages 130–139. Springer, 1923.

[75] William Feller. An introduction to probability and its applications, vol. ii. *Wiley, New York*, 1971.

[76] F Feroz, MP Hobson, E Cameron, and AN Pettitt. Importance nested sampling and the multinest algorithm. *arXiv preprint arXiv:1306.2144*, 2013.

[77] Pedro G Ferreira, Joao Magueijo, and Joseph Silk. Cumulants as non-gaussian qualifiers. *Physical Review D*, 56(8):4592, 1997.

[78] E. E. Flanagan. Sensitivity of the Laser Interferometer Gravitational Wave Observatory to a stochastic background, and its dependence on the detector orientations. *Physical Review D*, 48:2389–2407, September 1993.

[79] Eanna E Flanagan. Sensitivity of the laser interferometer gravitational wave observatory to a stochastic background, and its dependence on the detector orientations. *Physical Review D*, 48(6):2389, 1993.

[80] Éanna É Flanagan and Scott A Hughes. The basics of gravitational wave theory. *New Journal of Physics*, 7(1):204, 2005.

[81] A Ronald Gallant and Douglas W Nychka. Semi-nonparametric maximum likelihood estimation. *Econometrica: Journal of the Econometric Society*, pages 363–390, 1987.

[82] A Ronald Gallant and George Tauchen. Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. *Econometrica: Journal of the Econometric Society*, pages 1091–1120, 1989.

[83] E Gaztanaga, P Fosalba, and E Elizalde. Gravitational evolution of the large-scale probability density distribution: the edgeworth and gamma expansions. *The Astrophysical Journal*, 539(2):522, 2000.

[84] Ortwin E Gerhard. Line-of-sight velocity profiles in spherical galaxies: breaking the degeneracy between anisotropy and mass. *Monthly Notices of the Royal Astronomical Society*, 265(1):213–230, 1993.

[85] K Glazebrook, D Eisenstein, A Dey, and R Nichol. Dark energy and cosmic sound. *Dark Energy Task Force white paper (astro-ph/0507457)*, 2005.

[86] PC Gregory and AR Taylor. New highly variable radio source, possible counterpart of γ-ray source cg135+ 1. 1978.

[87] L. P. Grishchuk. Amplification of gravitational waves in an isotropic universe. *Soviet Journal of Experimental and Theoretical Physics*, 40:409, September 1975.

[88] L. P. Grishchuk. Relic gravitational waves and limits on inflation. *Physical Review D*, 48:3513–3516, October 1993.

[89] LP Grishchuk. Quantum effects in cosmology. *Classical and Quantum Gravity*, 10(12):2449, 1993.

[90] Alan H Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Physical Review D*, 23(2):347, 1981.

[91] Mark Hannam, Patricia Schmidt, Alejandro Bohé, Leïla Haegel, Sascha Husa, Frank Ohme, Geraint Pratten, and Michael Pürrer. Simple model of complete precessing black-hole-binary gravitational waveforms. *Physical Review Letters*, 113(15):151101, 2014.

[92] Donald R Herriott and Harry J Schulte. Folded optical delay lines. *Applied Optics*, 4(8):883–889, 1965.

[93] Jeremy S Heyl, Lars Hernquist, and David N Spergel. Structure of merger remnants. 4: Isophotal shapes. *The Astrophysical Journal*, 427:165–173, 1994.

[94] David Hilbert. Die grundlagen der physik.(erste mitteilung.). *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1915:395–408, 1915.

[95] S Hild, M Abernathy, F Acernese, P Amaro-Seoane, N Andersson, K Arun, F Barone, B Barr, M Barsuglia, M Beker, et al. Sensitivity studies for third-generation gravitational wave observatories. *Classical and Quantum Gravity*, 28(9):094013, 2011.

[96] Wayne Hu and Martin White. A cmb polarization primer. *New Astronomy*, 2(4):323–344, 1997.

[97] Edwin Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences*, 15(3):168–173, 1929.

[98] RA Hulse and JH Taylor. A deep sample of new pulsars and their spatial extent in the galaxy. *The Astrophysical Journal*, 201:L55–L59, 1975.

[99] Joseph Oscar Irwin. On the frequency distribution of the means of samples from a population having any law of frequency with finite moments, with special reference to pearson's type ii. *Biometrika*, pages 225–239, 1927.

[100] Mark G Jackson, Nicholas T Jones, and Joseph Polchinski. Collisions of cosmic f-and d-strings. *Journal of High Energy Physics*, 2005(10):013, 2005.

[101] Mark G Jackson and Xavier Siemens. Gravitational wave bursts from cosmic superstring reconnections. *Journal of High Energy Physics*, 2009(06):089, 2009.

[102] Piotr Jaranowski and Andrzej Królak. Gravitational-wave data analysis. formalism and sample applications: the gaussian case. *arXiv preprint arXiv:0711.1115*, 2007.

[103] Piotr Jaranowski and Andrzej Królak. *Analysis of gravitational-wave data*, volume 29. Cambridge University Press, 2009.

[104] Carlos M Jarque and Anil K Bera. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3):255–259, 1980.

[105] Eunhwa Jeong and George F Smoot. Search for cosmic strings in cosmic microwave background anisotropies. *The Astrophysical Journal*, 624(1):21, 2005.

[106] Nicholas Jones, Horace Stoica, and S-H Henry Tye. Brane interaction as the origin of inflation. *Journal of High Energy Physics*, 2002(07):051, 2002.

[107] L Ju, DG Blair, and C Zhao. Detection of gravitational waves. *Reports on Progress in Physics*, 63(9):1317, 2000.

[108] R Juszkiewicz, DH Weinberg, and P Chodorowski Amsterdamski. M. & bouchet f. 1995. *ApJ*, 442:39.

[109] Thomas WB Kibble. Topology of cosmic domains and strings. *Journal of Physics A: Mathematical and General*, 9(8):1387, 1976.

[110] Martin Kilbinger. Cosmology with cosmic shear observations: a review. *Reports on Progress in Physics*, 78(8):086901, 2015.

[111] Lev Kofman, Edmund Bertschinger, James M Gelb, Adi Nusser, and Avishai Dekel. Evolution of one-point distributions from gaussian initial fluctuations. *arXiv preprint astro-ph/9311028*, 1993.

[112] John Michael Kovac, EM Leitch, C Pryke, JE Carlstrom, NW Halverson, and WL Holzapfel. Detection of polarization in the cosmic microwave background using dasi. *Nature*, 420(6917):772–787, 2002.

[113] K Kuroda, LCGT collaboration, et al. Status of lcgt. *Classical and Quantum Gravity*, 27(8):084004, 2010.

[114] Lucien Le Cam. *Asymptotic methods in statistical decision theory*. Springer Science & Business Media, 2012.

[115] Andrew R Liddle and David H Lyth. *Cosmological inflation and large-scale structure*. Cambridge University Press, 2000.

[116] Hubert W Lilliefors. On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62(318):399–402, 1967.

[117] Alejandro Lopez and Katherine Freese. First test of high frequency gravity waves from inflation using advanced ligo. *Journal of Cosmology and Astroparticle Physics*, 2015(01):037, 2015.

[118] Duncan R Lorimer. Binary and millisecond pulsars. *Living Reviews in Relativity*, 11(8):21, 2008.

[119] G Losurdo. Advanced virgo sensitivity curve: cavity finesse and signal recycling tuning. Technical report, Virgo Internal Report: VIR024A07, 2007.

[120] Michele Maggiore. Gravitational wave experiments and early universe cosmology. *Physics Reports*, 331(6):283–367, 2000.

[121] Lionel Martellini and Tania Regimbau. Semiparametric approach to the detection of non-gaussian gravitational wave stochastic backgrounds. *Physical Review D*, 89(12):124009, 2014.

[122] Lionel Martellini and Tania Regimbau. Efficiency of the cross-correlation statistic for gravitational wave stochastic background signals with non-gaussian noise and heterogeneous detector sensitivities. *Physical Review D*, 92(10):104025, 2015.

[123] DV Martynov, ED Hall, BP Abbott, R Abbott, TD Abbott, C Adams, RX Adhikari, RA Anderson, SB Anderson, K Arai, et al. Sensitivity of the advanced ligo detectors at the beginning of gravitational wave astronomy. *Physical Review D*, 93(11):112004, 2016.

[124] David McClelland, Matthew Evans, Brian Lantz, Ian Martin, Volker Quetschke, and Roman Schnabel. Instrument science white paper 2015. 2015. Available at https://dcc.ligo.org/public/0120/T1500290/003/T1500290.pdf.

[125] Jessica McIver. Data quality studies of enhanced interferometric gravitational wave detectors. *Classical and Quantum Gravity*, 29(12):124010, 2012.

[126] Duncan Meacher, Michael Coughlin, Sean Morris, Tania Regimbau, Nelson Christensen, Shivaraj Kandhasamy, Vuk Mandic, Joseph D Romano, and Eric Thrane. Mock data and science challenge for detecting an astrophysical stochastic gravitational-wave background with advanced ligo and advanced virgo. *Physical Review D*, 92(6):063002, 2015.

[127] Ramon Miquel. Cosmology with type-ia supernovae. *Journal of Physics A: Mathematical and Theoretical*, 40(25):6743, 2007.

[128] Charles W Misner. The isotropy of the universe. *The Astrophysical Journal*, 151:431, 1968.

[129] R Moessner. Statistics of peculiar velocities from cosmic strings. *Monthly Notices of the Royal Astronomical Society*, 277(3):927–932, 1995.

[130] Alessandro Nagar, Thibault Damour, and Angelo Tartaglia. Binary black hole merger in the extreme-mass-ratio limit. *Classical and Quantum Gravity*, 24(12):S109, 2007.

[131] Jerzy Neyman and Egon S Pearson. On the problem of the most efficient tests of statistical hypotheses. In *Breakthroughs in Statistics*, pages 73–108. Springer, 1992.

[132] Christian D Ott. The gravitational-wave signature of core-collapse supernovae. *Classical and Quantum Gravity*, 26(6):063001, 2009.

[133] Christian D Ott, Adam Burrows, Luc Dessart, and Eli Livne. A new mechanism for gravitational-wave emission in core-collapse supernovae. *Physical Review Letters*, 96(20):201102, 2006.

[134] Gamze Özel. On the moment characteristics for the univariate compound poisson and bivariate compound poisson processes with applications. *Revista Colombiana de Estadística*, 36(1):59–77, 2013.

[135] Phillip James Edwin Peebles. *Principles of physical cosmology*. Princeton University Press, 1993.

[136] Arno A Penzias and Robert Woodrow Wilson. A measurement of excess antenna temperature at 4080 mc/s. *The Astrophysical Journal*, 142:419–421, 1965.

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

[138] PC Peters and Jon Mathews. Gravitational radiation from point masses in a keplerian orbit. *Physical Review*, 131(1):435, 1963.

[139] Philip Carl Peters. Gravitational radiation and the motion of two point masses. *Physical Review*, 136(4B):B1224, 1964.

[140] FAE Pirani. On the physical significance of the riemann tensor. *Acta Physica Polonica*, 15:389–405, 1956.

[141] Frans Pretorius. Evolution of binary black-hole spacetimes. *Physical Review Letters*, 95(12):121101, 2005.

[142] M Punturo et al. The einstein telescope: a third-generation gravitational wave observatory. *Classical and Quantum Gravity*, 27(19):194002, 2010.

[143] T. Regimbau, M. Evans, N. Christensen, E. Katsavounidis, B. Sathyaprakash, and S. Vitale. Digging deeper: Observing primordial gravitational waves below the binary-black-hole-produced stochastic background. *Phys. Rev. Lett.*, 118:151105, Apr 2017.

[144] T. Regimbau, S. Giampanis, X. Siemens, and V. Mandic. Stochastic background from cosmic (super)strings: Popcorn-like and (Gaussian) continuous regimes. *Physical Review D*, 85(6):066001, March 2012.

[145] Tania Regimbau. The astrophysical gravitational wave stochastic background. *Research in Astronomy and Astrophysics*, 11(4):369, 2011.

[146] P Reig, A Nersesian, A Zezas, L Gkouvelis, and MJ Coe. Long-term optical variability of high-mass x-ray binaries-ii. spectroscopy. *Astronomy & Astrophysics*, 590:A122, 2016.

[147] Adam G Riess, Alexei V Filippenko, Peter Challis, Alejandro Clocchiatti, Alan Diercks, Peter M Garnavich, Ron L Gilliland, Craig J Hogan, Saurabh Jha, Robert P Kirshner, et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, 116(3):1009, 1998.

[148] Hans-Walter Rix, Marijn Franx, David Fisher, and Garth Illingworth. Ngc 4550-a laboratory for testing galaxy formation. *The Astrophysical Journal*, 400:L5–L8, 1992.

[149] Joseph D Romano and Neil J Cornish. Detection methods for stochastic gravitational-wave backgrounds: A unified treatment. *arXiv preprint arXiv:1608.06889*, 2016.

[150] Pablo A Rosado. Gravitational wave background from binary systems. *Physical Review D*, 84(8):084004, 2011.

[151] Pablo A Rosado. Gravitational wave background from rotating neutron stars. *Physical Review D*, 86(10):104007, 2012.

[152] Christian Röver. Student-t based filter for robust signal detection. *Physical Review D*, 84(12):122004, 2011.

[153] Peter R Saulson. Physics of gravitational wave detection: resonant and interferometric detectors. *XXVI SLAC Summer Institute on Particle Physics*, 1998.

[154] Robert J Scherrer and Edmund Bertschinger. Statistics of primordial density perturbations from discrete seed masses. *The Astrophysical Journal*, 381:349–360, 1991.

[155] L Schnupp, W Winkler, K Maischberger, A Rudiger, and R Schilling. Reduction of noise due to scattered light in gravitational wave antennae by modulating the phase of the laser light. *Journal of Physics E: Scientific Instruments*, 18(6):482, 1985.

[156] Anand Sengupta. The sensitivity of the advanced ligo detectors at the beginning of gravitational wave astronomy. 2016.

[157] N. Seto. Non-Gaussianity analysis of a gravitational wave background made by short-duration burst signals. *Physical Review D*, 80(4):043003, August 2009.

[158] D Shoemaker, R Schilling, L Schnupp, W Winkler, K Maischberger, and A Rüdiger. Noise behavior of the garching 30-meter prototype gravitational-wave detector. *Physical Review D*, 38(2):423, 1988.

[159] Xavier Siemens, Jolien Creighton, Irit Maor, Saikat Ray Majumder, Kipp Cannon, and Jocelyn Read. Gravitational wave bursts from cosmic (super) strings: Quantitative analysis and constraints. *Physical Review D*, 73(10):105001, 2006.

[160] Xavier Siemens, Vuk Mandic, and Jolien Creighton. Gravitational-wave stochastic background from cosmic strings. *Physical Review Letters*, 98(11):111101, 2007.

[161] Nickolay Smirnov. Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, 19(2):279–281, 1948.

[162] David Nathaniel Spergel, R Bean, O Doré, MR Nolta, CL Bennett, J Dunkley, G Hinshaw, N Jarosik, E Komatsu, L Page, et al. Three-year wilkinson microwave anisotropy probe (wmap) observations: implications for cosmology. *The Astrophysical Journal Supplement Series*, 170(2):377, 2007.

[163] A. A. Starobinskii. Spectrum of relict gravitational radiation and the early state of the universe. *ZhETF Pisma Redaktsiiu*, 30:719–723, December 1979.

[164] Joseph H Taylor and Joel M Weisberg. A new test of general relativity-gravitational radiation and the binary pulsar psr 1913+ 16. *The Astrophysical Journal*, 253:908–920, 1982.

[165] Joseph H Taylor and Joel M Weisberg. Further experimental tests of relativistic gravity using the binary pulsar psr 1913+ 16. *The Astrophysical Journal*, 345:434–450, 1989.

[166] Advanced LIGO Team et al. Advanced ligo reference design. *LIGO Document*, 60056, 2007.

[167] KS Thorne. In sw hawking and w. israel, editors. *Three Hundred Years of Gravitation*, pages 330–458.

[168] E. Thrane. Measuring the non-Gaussian stochastic gravitational-wave background: A method for realistic interferometer data. *Physical Review D*, 87(4):043009, February 2013.

[169] Eric Thrane and Joseph D Romano. Sensitivity curves for searches for gravitational-wave backgrounds. *Physical Review D*, 88(12):124032, 2013.

[170] Greg Ushomirsky, Curt Cutler, and Lars Bildsten. Deformations of accreting neutron star crusts and gravitational wave emission. *Monthly Notices of the Royal Astronomical Society*, 319(3):902–932, 2000.

[171] Roeland P Van Der Marel and Marijn Franx. A new method for the identification of non-gaussian line profiles in elliptical galaxies. *The Astrophysical Journal*, 407:525–539, 1993.

[172] Alexander Vilenkin and E Paul S Shellard. *Cosmic strings and other topological defects*. Cambridge University Press, 2000.

[173] Joseph Weber. Evidence for discovery of gravitational radiation. *Physical Review Letters*, 22(24):1320, 1969.

[174] Steven Weinberg. *Gravitation and cosmology: principles and applications of the general theory of relativity*, volume 1. Wiley New York, 1972.

[175] Steven Weinberg. *Cosmology*. Oxford University Press, 2008.

[176] Joel M Weisberg and Yuping Huang. Relativistic measurements from timing the binary pulsar psr b1913+ 16. *arXiv preprint arXiv:1606.02744*, 2016.

[177] Joel M Weisberg, David J Nice, and Joseph H Taylor. Timing measurements of the relativistic binary pulsar psr b1913+ 16. *The Astrophysical Journal*, 722(2):1030, 2010.

[178] R Weiss. Quarterly progress report. *MIT Research Lab of Electronics*, 105:54, 1972.

[179] Edward L Wright. Theoretical overview of cosmic microwave background anisotropy. *arXiv preprint astro-ph/0305591*, 2003.

[180] Chengjiang Wu, Vuk Mandic, and Tania Regimbau. Accessibility of the gravitational-wave background due to binary coalescences to second and third generation gravitational-wave detectors. *Physical Review D*, 85(10):104024, 2012.

[181] Takahiro Yamamoto, Kazuhiro Hayama, Shuhei Mano, Yousuke Itoh, and Nobuyuki Kanda. Characterization of non-gaussianity in gravitational wave detector noise. *Physical Review D*, 93(8):082005, 2016.

[182] Xing-Jiang Zhu, Eric Howell, Tania Regimbau, David Blair, and Zong-Hong Zhu. Stochastic gravitational wave background from coalescing binary black holes. *The Astrophysical Journal*, 739(2):86, 2011.

[183] Mark Zimmermann and Eugene Szedenits Jr. Gravitational waves from rotating and precessing rigid bodies: Simple models and applications to pulsars. *Physical Review D*, 20(2):351, 1979.