Detection methods for stochastic gravitational-wave backgrounds: A unified treatment

Joseph D. Romano
Department of Physics
University of Texas Rio Grande Valley, Brownsville, TX 78520
email: joseph.romano@utrgv.edu

Neil J. Cornish
Department of Physics
Montana State University, Bozeman, MT 59717
email: cornish@physics.montana.edu

Abstract

We review detection methods that are currently in use or have been proposed to search for a stochastic background of gravitational radiation. We consider both Bayesian and frequentist searches using ground-based and space-based laser interferometers, spacecraft Doppler tracking, and pulsar timing arrays; and we allow for anisotropy, non-Gaussianity, and non-standard polarization states. Our focus is on relevant data analysis issues, and not on the particular astrophysical or early Universe sources that might give rise to such backgrounds. We provide a unified treatment of these searches at the level of detector response functions, detection sensitivity curves, and, more generally, at the level of the likelihood function, since the choice of signal and noise models and prior probability distributions are actually what define the search. Pedagogical examples are given whenever possible to compare and contrast different approaches. We have tried to make the article as self-contained as possible, targeting graduate students and new researchers looking to enter this field.

Contents

1 Introduction

> There's two possible outcomes: if the result confirms the hypothesis, then you've made a discovery. If the result is contrary to the hypothesis, then you've made a discovery. *Enrico Fermi*

It is an exciting time for the field of gravitational-wave astronomy. The observation, on September 14th, 2015, of gravitational waves from the inspiral and merger of a pair of black holes [12] has opened a radically new way of observing the Universe. The event, denoted GW150914, was observed simultaneously by the two detectors of the Laser Interferometer Gravitational-wave Observatory (LIGO) [3]. [LIGO consists of two 4 km-long laser interferometers, one located in Hanford, Washington, the other in Livingston, LA.] The merger event that produced the gravitational waves occured in a distant galaxy roughly 1.3 billion light years from Earth. The initial masses of the two black holes were estimated to be 36^{+5}_{-4} $M_\odot$ and 29^{+4}_{-4} $M_\odot$, and that of the post-merger black hole as 62^{+4}_{-4} $M_\odot$ [171]. The difference between the initial and final masses corresponds to $3.0^{+0.5}_{-0.5}$ $M_\odot c^2$ of energy radiated in gravitational waves, with a peak luminosity of *more than ten times the combined luminosity of all the stars in all the galaxies in the visible universe*! The fact that this event was observed *only* in gravitational waves—and not in electromagnetic waves—illustrates the complementarity and potential for new discoveries that comes with the opening of the gravitational-wave window onto the universe.

GW150914 is just the first of many gravitational-wave signals that we expect to observe over the next several years. Indeed, roughly three months after the detection of GW150914, a second event, GW151226, was observed by the two LIGO detectors [11]. This event also involved the inspiral and merger of a pair of stellar mass black holes, with initial component masses $14.2^{+8.3}_{-3.7}$ $M_\odot$ and $7.5^{+2.3}_{-2.3}$ $M_\odot$, and a final black hole mass of $20.8^{+6.1}_{-1.7}$ $M_\odot$. The source was at a distance of roughly 1.4 billion light-years from Earth, comparable to that of GW150914. Advanced LIGO will continue interleaving observation runs and commissioning activities to reach design sensivity around 2020 [3], which will allow detections of signals like GW150914 and GW1512126 with more than three times the signal-to-noise ratio than was observed for GW150914 (which was 24). In addition, the Advanced Virgo detector [15] (a 3 km-long laser interferometer in Cascina, Italy) and KAGRA [33] (a 3 km-long cryogenic laser interferometer in Kamioka mine in Japan) should both be taking data by the end of 2016. There are also plans for a third LIGO detector in India [92]. A global network of detectors such as this will allow for much improved position reconstruction and parameter estimation of the sources [13].

1.1 Motivation and context

GW150914 and GW151226 were single events—binary black hole mergers that had relatively large matched-filter signal-to-noise ratios (24 and 13, respectively) in the two LIGO detectors [12, 11]. But for every loud event like GW150914 or GW151226, there are many more quiet events that are too distant to be individually detected (the associated signal-to-noise ratios are too low). The total rate of merger events from the population of stellar-mass binary black holes of which GW150914 and GW151226 are members can be estimated by taking the local rate estimate 9–240 $\mathrm{Gpc}^{-3}\,\mathrm{yr}^{-1}$ [170] and multiplying by the comoving volume out to some large redshift, e.g., $z \sim 6$. This yields a total rate of binary black hole mergers between ~ 1 per minute and a few per hour. Since the duration of each merger signal at Earth is of order a few tenths of a second to ~ 1 second, the *duty cycle* (defined as one minus the fraction of time that the signal is "off" in the data) is $\ll 1$. This means that the combined signal from such a population of binary black holes will be "popcorn-like", with the majority of the individual signals being too weak to individually detect. Since the arrival times of the merger signals are randomly-distributed, the combined signal from the population of binary black holes is itself random—it is an example of a *stochastic background* of gravitational radiation.

More generally, a stochastic background of gravitational radiation is *any* random gravitational-wave signal produced by a large number of weak, independent, and unresolved sources. The background doesn't have to be popcorn-like, like the expected signal from the population of binary black holes which gave rise to GW150914 and GW151226. It can be composed of individual deterministic signals that overlap in time (or in frequency) producing a "confusion" noise analogous to conversations at a cocktail party. Such a confusion background (in this case, it will actually be a *foreground*) is produced by the galactic population of compact white dwarf binaries. Alternatively, the signal can be *intrinsically* random, associated with stochastic processes in the early Universe.

The focus of this review article is on data analysis strategies (i.e., detection methods) that can be used to detect and ultimately characterize a stochastic gravitational-wave background. To introduce this topic and to set the stage for the more detailed discussions to follow in later sections, we ask (and start to answer) the following questions:

1.1.1 Why do we care about detecting a stochastic background?

Detecting a stochastic background of gravitational radiation can provide information about astrophysical source populations and processes in the very early Universe, which are inaccessible by any other means. For example, electromagnetic radiation cannot provide a picture of the Universe any earlier than the time of last of scattering (roughly 400,000 years after the Big Bang). Gravitational waves, on the other hand, can give us information all the way back to the onset of inflation, a mere $\sim 10^{-32}$ s after the Big Bang. (See [115] for a detailed discussion of both cosmological and astrophysical sources of a stochastic gravitational-wave background.)

1.1.2 Why is detection challenging?

Stochastic signals are effectively another source of noise in a single detector. So the fundamental problem is how to distinguish between gravitational-wave "noise" and instrumental noise. It turns out that there are several ways to do this, as we will discuss in the later sections of this article.

1.1.3 What detection methods can one use?

Cross-correlation methods can be used whenever one has multiple detectors that respond to the common gravitational-wave background. For single detector analyses (e.g., for eLISA), one needs to take advantage of null combinations of the data (which acts an instrument noise monitor) or use instrument noise modeling to try to distinguish the gravitational-wave signal from instrument noise. Over the past 15 years or so, the number of detection methods for stochastic backgrounds has increased considerably. So now, in addition to the standard cross-correlation search for a "vanilla" (Gaussian-stationary, unpolarized, isotropic) background, one can search for non-Gaussian backgrounds, anisotropic backgrounds, circularly-polarized backgrounds, and backgrounds with polarization components predicted by (non-general-relativity) modified theories of gravity. These searches are discussed in Sections 7 and 8.

Table 1 summarizes the basic properties of various analysis methods that have been used (or proposed) for stochastic background searches. Despite apparent differences, *all*

EARLY ANALYSES (before 2000)	MORE RECENT ANALYSES
used frequentist statistics	use both frequentist and Bayesian inference
used cross-correlation methods	use cross-correlation methods and stochastic templates; use null channels or knowledge about instrumental noise when cross-correlation is not available
assumed Gaussian noise	have allowed non-Gaussian noise
assumed stationary, Gaussian, unpolarized, and isotropic gravitational-wave backgrounds	have allowed non-Gaussian, polarized, and anisotropic gravitational-wave backgrounds
were done primarily in the context of ground-based detectors (e.g., resonant bars and LIGO-like interferometers) where the small-antenna (i.e., long-wavelength) approximation was valid	have been done in the context of space-based detectors (e.g., spacecraft tracking, LISA) and pulsar timing arrays for which the small-antenna approximation is not valid

Table 1: Overview of analysis methods for stochastic gravitational-wave backgrounds. The number and flexibility of the methods have increased considerably since the year 2000.

analyses use a likelihood function, e.g., for defining frequentist statistics or for calculating posterior distributions for Bayesian inference (as will be described in more detail in Section 3), and take advantage of cross-correlations if multiple detectors are available (as will be described in more detail in Section 4).

1.1.4 What are the prospects for detection?

The prospects for detection depend on the source of the background (i.e., astrophysical or cosmological) and the type of detector being used. For example, a space-based interferometer like eLISA is *guaranteed* to detect the gravitational-wave confusion noise produced by the galactic population of compact white dwarf binaries. Pulsar timing arrays, on the other hand, are likely to see the confusion noise from supermassive black hole binaries (SMBHBs) at the centers of merging galaxies. Detection sensitivity curves are a very convenient way of comparing theoretical predictions of source strengths to the sensivity levels of the various detectors (see Section 10 for details).

1.2 Searches across the gravitational-wave spectrum

The frequency band of ground-based laser interferometers like LIGO, Virgo, and KAGRA is between ~ 10 Hz and a few kHz (gravity gradient and seismic noise are the limiting noise sources below 10 Hz, and photon shot noise above a couple of kHz). Outside this band there are several other experiments—both currently operating and planned—that should also detect gravitational waves in the not too distant future. An illustration of the gravitational-wave spectrum, together with potential sources and relevant detectors, is shown in Figure 1. We highlight a few of these experiments below.

1.2.1 Cosmic microwave background experiments

At the extreme low-frequency end of the spectrum, corresponding to gravitational-wave periods of order the age of the Universe, the Planck satellite [187] and other cosmic microwave background (CMB) experiments, e.g., BICEP and Keck [184] are looking for evidence of relic gravitational waves from the Big Bang in the B-mode component of CMB polarization maps [98, 90, 39]. In 2014, BICEP2 announced the detection of relic gravitational waves [18], but it was later shown that the observed B-mode signal was due to contamination by intervening dust in the galaxy [67, 125]. So at present, these experiments have been able to only *constrain* (i.e., set upper limits on) the amount of gravitational waves in the very early Universe [39]. But these constraints severely limit the possibility of detecting the relic gravitational-wave background with any of the higher-frequency detection methods, unless its spectrum increases with frequency. [Note that standard models of inflation predict a relic background whose energy density is almost constant in frequency, leading to a strain spectral density that decreases with frequency.] Needless to say, the detection of a primordial gravitational-wave background is a "holy grail" of gravitational-wave astronomy.

1.2.2 Pulsar timing arrays

At frequencies between $\sim 10^{-9}$ Hz and 10^{-7} Hz, corresponding to gravitational-wave periods of order decades to years, pulsar timing arrays (PTAs) can be used to search for gravitational waves. This is done by carefully monitoring the arrival times of radio pulses from an array of galactic millisecond pulsars, looking for *correlated* modulations in the arrival times induced by a passing gravitational wave [57, 82]. The most-likely

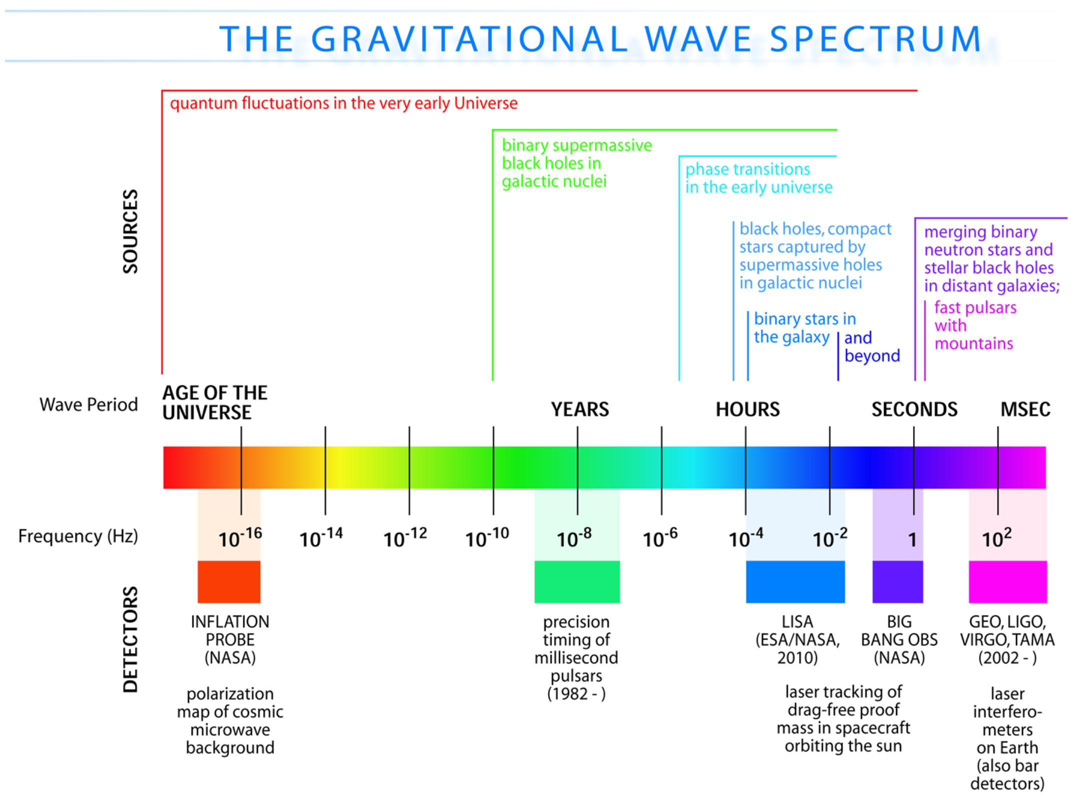


Figure 1: Gravitational-wave spectrum, together with potential sources and relevant detectors. Image credit: Institute of Gravitational Research/ University of Glasgow.

gravitational-wave source for PTAs is a gravitational-wave background formed from the incoherent superposition of signals produced by the inspirals and mergers of SMBHBs in the centers of distant galaxies [94]. These searches continue to improve their sensitivity by upgrading instrument back-ends and discovering more millisecond pulsars that can be added to the array. These improvements have led to more constraining upper limits on the amplitude of the gravitational-wave background [157, 31], with a detection being likely before the end of this decade [159, 168].

1.2.3 Space-based interferometers

At frequencies between $\sim 10^{-4}$ Hz and 10^{-2} Hz, corresponding to gravitational-wave periods of order hours to minutes, proposed space-based interferometers like LISA or eLISA [188] can search for gravitational waves from a wide variety of sources [72]. These include: (i) inspirals and mergers of SMBHBs with masses of order 10^6 $M_\odot$, (ii) captures of compact stellar-mass objects around supermassive black holes, and (iii) the stochastic confusion noise produced by compact white-dwarf binaries in our galaxy. The basic space-based interferometer configuration consists of three satellites (each housing two lasers, two telescopes, and two test masses) that fly in an equilateral-triangle formation, with arm lengths of order one-million km. eLISA is currently being considered by the European Space Agency (ESA) as the 3rd large mission in its Cosmic Vision Program [185]. The earliest launch date for eLISA would be around 2030. A technology-demonstration mission, called LISA Pathfinder [186], was launched in December 2015, meeting or exceeding all technology requirements [29].

1.3 Goal of this article

Starting with the pioneering work of Grishchuk [78], Detweiler [57], Hellings and Downs [82], and Michelson [119], detection methods for gravitational-wave backgrounds have increased in scope and sophistication over the years, with several new developments occuring rather recently. As mentioned above, we have search methods now that target different properties of the background (e.g., isotropic or anisotropic, Gaussian or non-Gaussian, polarized or unpolarized, etc.). These searches are necessarily implemented differently for different detectors, since, for example, ground-based detectors like LIGO and Virgo operate in the *small-antenna* (or *long-wavelength*) limit, while pulsar timing arrays operate in the *short-wavelength* limit. Moreover, each of these searches can be formulated in terms of either Bayesian or frequentist statistics. *The goal of this review article is to discuss these different detection methods from a perspective that attempts to* unify *the different treatments, emphasizing the similarities that exist when viewed from this broader perspective.*

1.4 Unification

The extensive literature describing stochastic background analyses leaves the reader with the impression that highly specialized techniques are needed for ground, space, and pulsar timing observations. Moreover, reviews of gravitational-wave data analysis leave the impression that the analysis of stochastic signals is somehow fundamentally different from that of any other signal type. Both of these impressions are misleading. The apparent differences are due to differences in terminology and perspective. By adopting a common analysis framework and notation, we are able to present a *unified* treatment of gravitational-wave data analysis across source classes and observation techniques.

We will provide a unified treatment of the various methods at the level of detector response functions, detection sensitivity curves, and, more generally, at the level of the likelihood function, since the choice of signal and noise models and prior probability distributions are actually what define the search. The same photon time-of-flight calculation underpins the detector response functions, and the choice of prior for the gravitational-wave template defines the search. A *matched-filter* search for binary mergers and a *cross-correlation* search for stochastic signals are both derived from the same likelihood function, the difference being that the former uses a parameterized, deterministic template, while the latter uses a stochastic template. Hopefully, by the end of this article, the reader will see that the plethora of searches for different types of backgrounds, using different types of detectors, and using different statistical inference frameworks are not all that different after all.

1.5 Outline

The rest of the article is organized as follows: We begin in Section 2 by specifying the quantities that one uses to characterize a stochastic gravitational-wave background. In Section 3, we give an overview of statistical inference by comparing and contrasting how the Bayesian and frequentist formalisms address issues related to hypothesis testing, model selection, setting upper limits, parameter estimation, etc. We then illustrate these concepts in the context of a very simple toy problem. In Section 4, we introduce the key concept of correlation, which forms the basis for the majority of detection methods used for gravitational-wave backgrounds, and show how these techniques arise naturally from the standard template-based approach. We derive the frequentist cross-correlation statistic for a simple example. We also describe how a null channel is useful when correlation methods are not possible.

In Section 5, we go into more detail regarding the different types of detectors. In particular, we calculate single-detector response functions and the associated antenna patterns for ground-based and space-based laser interferometers, spacecraft Doppler tracking, and pulsar timing measurements. (We do not discuss resonant bar detectors or CMB-based detection methods in this review article. However, current bounds from CMB observations will be reviewed in Section 10.) By correlating the outputs of two such detectors, we obtain expressions for the correlation coefficient (or *overlap reduction function*) for a Gaussian-stationary, unpolarized, isotropic background as a function of the separation and orientation of the two detectors. In Section 6, we discuss optimal filtering. Section 7 extends the analysis of the previous sections to *anisotropic* backgrounds. Here we describe several different analyses that produce maps of the gravitational-wave sky: (i) a frequentist gravitational-wave radiometer search, which is optimal for point sources, (ii) searches that decompose the gravitational-wave power on the sky in terms of spherical harmonics, and (iii) a phase-coherent search that can map both the amplitude and phase of a gravitational-wave background at each location on the sky. In Section 8, we discuss searches for: (i) non-Gaussian backgrounds, (ii) circularly-polarized backgrounds, and (iii) backgrounds having non-standard (i.e., non-general-relativity) polarization modes. We also briefly describe extensions of the cross-correlation search method to look for *non-stochastic-background-type* signals—in particular, long-duration unmodelled transients and continuous (nearly-monochromatic) gravitational-wave signals from sources like Sco X-1.

In Section 9, we discuss real-world complications introduced by irregular sampling, non-stationary and non-Gaussian detector noise, and correlated environmental noise (e.g., Schumann resonances). We also describe what one can do if one has only a single detector, as is the case for eLISA. Finally, we conclude in Section 10 by discussing prospects for detection, including detection sensitivity curves and current observational results.

We also include several appendices: In Appendix A we discuss different polarization basis tensors, and a Stokes' parameter characterization of gravitational-waves. In Appendices B and C, we summarize some standard statistical results for a Gaussian random variable, and then discuss how to define and test for non-stationarity and non-Gaussianity. Appendices D, E, F are adapted from [71], with details regarding spin-weighted scalar, vector, and tensor spherical harmonics. Finally, Appendix G gives a "Rosetta stone" for translating back and forth between different response function conventions for gravitational-wave backgrounds.

2 Characterizing a stochastic gravitational-wave background

> When you can measure what you are speaking about, and express it in numbers, you know something about it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarely, in your thoughts advanced to the stage of science. *William Thomson, Baron Kelvin of Largs*

In this section, we define several key quantities (e.g., fractional energy density spectrum, characteristic strain, distribution of gravitational-wave power on the sky), which are used to characterize a stochastic background of gravitational radiation. The definitions are appropriate for both isotropic and anisotropic backgrounds. Our approach is similar to that found in [26] for isotropic backgrounds and for the standard polarization basis. For the plane-wave decomposition in terms of tensor spherical harmonics, we follow [69, 71]. Detailed derivations can be found in those papers.

2.1 When is a gravitational-wave signal stochastic?

The standard "textbook" definition of a stochastic background of gravitational radiation is *a random gravitational-wave signal produced by a large number of weak, independent, and unresolved sources*. To say that it is random means that it can be characterized only statistically, in terms of expectation values of the field variables or, equivalently, in terms of the Fourier components of a plane-wave expansion of the metric perturbations (Section 2.3.1). If the number of independent sources is sufficiently large, the background will be Gaussian by the central limit theorem. Knowledge of the first two moments of the distribution will then suffice to determine all higher-order moments (Appendix B). For non-Gaussian backgrounds, third and/or higher-order moments will also be needed.

Although there is general agreement with the above definition, there has been some confusion and disagreement in the literature [143, 141, 140, 139] regarding some of the defining properties of a stochastic background. This is because terms like *weak* and *unresolved* depend on details of the observation (e.g., the sensitivity of the detector, the total observation time, etc.), which are not intrinsic properties of the background. So the answer to the question "When is a gravitational-wave signal stochastic?" is not as simple or obvious as it might initially seem.

In [44], we addressed this question in the context of searches for gravitational-wave backgrounds produced by a population of astrophysical sources. We found that it is best to give *operational* definitions for these properties, framed in the context of Bayesian inference. We will discuss Bayesian inference in more detail in Section 3, but for now the most important thing to know is that by using Bayesian inference we can calculate the probabilities of different signal-plus-noise models, given the observed data. The signal-plus-noise model with the largest probability is the preferred model, i.e., the one that is most consistent with the data. This is the essence of Bayesian model selection.

So we define a signal to be *stochastic* if a Bayesian model selection calculation prefers a stochastic signal model over any deterministic signal model. We also define a signal to be *resolvable* if it can be decomposed into *separate* (e.g., non-overlapping in either time or frequency) and *individually detectable* signals, again in a Bayesian model selection sense.[1] If the background is associated with the superposition of signals from many astrophysical sources—as we expect for the population of binary black holes which gave rise to GW150914 and GW151226—then we should *subtract out* any bright deterministic signals that standout above the lower-amplitude background, leaving behind a residual non-deterministic signal whose statistical properties we would like to determine. In the context of Bayesian inference, this 'subtraction' is done by allowing *hybrid* signal models, which consist of both parametrized deterministic signals and non-deterministic backgrounds. By using such hybrid models we can investigate the statistical properties of the residual background without the influence of the resolvable signals.

We will return to these ideas in Section 8.1, when we discuss searches for non-Gaussian backgrounds in more detail.

2.2 Plane-wave expansions

Gravitational waves are time-varying perturbations to the space-time metric, which propagate at the speed of light. In transverse-traceless coordinates, the metric perturbations $h_{ab}(t,\vec{x})$ corresponding to a gravitational-wave background can be written as a superposition of sinusoidal plane waves having frequency f, and coming from different directions $\hat{n}$ on the sky:[2]

$$h_{ab}(t,\vec{x}) = \int_{-\infty}^{\infty} df \int d^2\Omega_{\hat{n}}\, h_{ab}(f,\hat{n}) e^{i2\pi f(t+\hat{n}\cdot\vec{x}/c)} \tag{2.1}$$

For a stochastic background, the metric perturbations $h_{ab}(t,\vec{x})$ and hence the Fourier coefficients $h_{ab}(f,\hat{n})$ are random variables, whose probability distributions define the statistical properties of the background.

2.2.1 Polarization basis

Typically, one expands the Fourier coefficients $h_{ab}(f,\hat{n})$ in terms of the standard $+$ and $\times$ polarization tensors:

$$h_{ab}(f,\hat{n}) = h_+(f,\hat{n}) e^+_{ab}(\hat{n}) + h_\times(f,\hat{n}) e^\times_{ab}(\hat{n})\,, \tag{2.2}$$

where

$$\begin{aligned} e^+_{ab}(\hat{n}) &= \hat{l}_a\hat{l}_b - \hat{m}_a\hat{m}_b\,,\\ e^\times_{ab}(\hat{n}) &= \hat{l}_a\hat{m}_b + \hat{m}_a\hat{l}_b\,, \end{aligned} \tag{2.3}$$

and $\hat{l}$, $\hat{m}$ are the standard angular unit vectors tangent to the sphere:

$$\begin{aligned} \hat{n} &= \sin\theta\cos\phi\,\hat{x} + \sin\theta\sin\phi\,\hat{y} + \cos\theta\,\hat{z} \equiv \hat{r}\,,\\ \hat{l} &= \cos\theta\cos\phi\,\hat{x} + \cos\theta\sin\phi\,\hat{y} - \sin\theta\,\hat{z} \equiv \hat{\theta}\,,\\ \hat{m} &= -\sin\phi\,\hat{x} + \cos\phi\,\hat{y} \equiv \hat{\phi}\,. \end{aligned} \tag{2.4}$$

[1]Signals may be separable even when overlapping in time and frequency if the detector has good sky resolution, or if the signals have additional complexities due to effects such as orbital evolution and precession.

[2]The gravitational-wave propagation direction, which we will denote by $\hat{k}$, is given by $\hat{k} = -\hat{n}$.

(See Figure 2.) Searches for stochastic backgrounds having alternative polarization modes, as predicted by modified (metric) theories of gravity, will be discussed in Section 8.3.

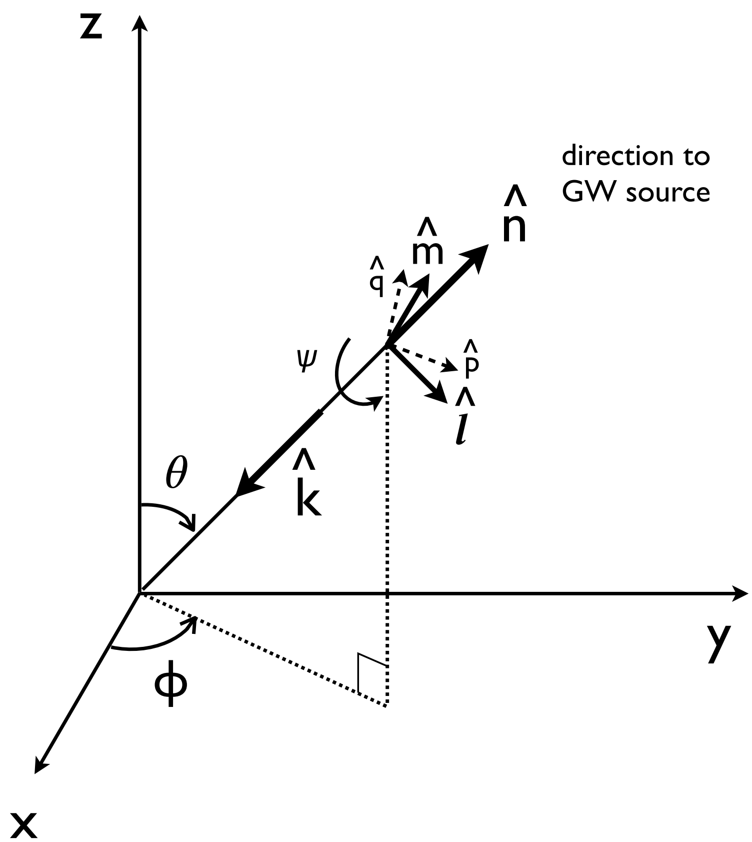


Figure 2: Our convention for the unit vectors $\{\hat{n}, \hat{l}, \hat{m}\}$ in terms of which the polarization basis tensors $e^{+}_{ab}(\hat{n})$ and $e^{\times}_{ab}(\hat{n})$ are defined. The unit vector $\hat{n}$ points in the direction of the gravitational-wave source (the gravitational wave propagates in direction $\hat{k} = -\hat{n}$); $\hat{l} = \hat{\theta}$ and $\hat{m} = \hat{\phi}$ are two unit vectors that lie in the plane perpendicular to $\hat{n}$. Another choice for the polarization basis tensors, defined in terms of the 'rotated' unit vectors $\hat{p}$ and $\hat{q}$, is given in Appendix A.

2.2.2 Tensor spherical harmonic basis

It is also possible to expand the Fourier coefficients $h_{ab}(f, \hat{n})$ in terms of the *gradient* and *curl* tensor spherical harmonics [69]:

$$h_{ab}(f, \hat{n}) = \sum_{l=2}^{\infty} \sum_{m=-l}^{l} \left[a^{G}_{(lm)}(f) Y^{G}_{(lm)ab}(\hat{n}) + a^{C}_{(lm)}(f) Y^{C}_{(lm)ab}(\hat{n}) \right] , \tag{2.5}$$

where

$$\begin{aligned} Y^G_{(lm)ab} &= {}^{(2)}N_l \left(Y_{(lm);ab} - \frac{1}{2} g_{ab} Y_{(lm);c}{}^{c} \right) , \\ Y^C_{(lm)ab} &= \frac{{}^{(2)}N_l}{2} \left(Y_{(lm);ac}\epsilon^c{}_b + Y_{(lm);bc}\epsilon^c{}_a \right) . \end{aligned} \tag{2.6}$$

In the above expressions, a semi-colon denotes covariant differentiation, g_{ab} is the metric tensor on the sphere, and ϵ_{ab} is the Levi-Civita anti-symmetric tensor. In standard spherical coordinates (θ, ϕ),

$$g_{ab} = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{pmatrix} , \qquad \epsilon_{ab} = \sqrt{g} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} , \qquad \sqrt{g} = \sin\theta . \tag{2.7}$$

The normalization constant

$${}^{(2)}N_l = \sqrt{\frac{2(l-2)!}{(l+2)!}} , \tag{2.8}$$

was chosen so that $\{Y^G_{(lm)ab}(\hat{n}), Y^C_{(lm)ab}(\hat{n})\}$ forms a set of orthonormal functions (with respect to the multipole indices l and m) on the 2-sphere. Appendix F contains additional details regarding gradient and curl spherical harmonics.

NOTE: we have adopted the notational convention used in the CMB literature, e.g., [98], by putting parentheses around the lm indices to distinguish them from the spatial tensor indices a, b, etc. In addition, summations over l and m start at $l = 2$, and not $l = 0$ as would be the case for the expansion of a scalar field on the 2-sphere in terms of ordinary (i.e., undifferentiated) spherical harmonics. In what follows, we will use $\sum_{(lm)}$ as shorthand notation for $\sum_{l=2}^{\infty}\sum_{m=-l}^{l}$ unless indicated otherwise.

2.2.3 Relating the two expansions

The gradient and curl spherical harmonics have been used extensively in the CMB community for decomposing CMB-polarization maps in terms of E-modes and B-modes (corresponding to the gradient and curl spherical harmonics). The most relevant property of the gradient and curl spherical harmonics is that they transform like combinations of spin-weight ± 2 fields with respect to rotations of an orthonomal basis at points on the 2-sphere. Explicitly,

$$Y^G_{(lm)ab}(\hat{n}) \pm i Y^C_{(lm)ab}(\hat{n}) = \frac{1}{\sqrt{2}} \left(e^+_{ab}(\hat{n}) \pm i e^{\times}_{ab}(\hat{n}) \right) {}_{\mp 2}Y_{lm}(\hat{n}) , \tag{2.9}$$

where ${}_{\pm 2}Y_{lm}(\hat{n})$ are the spin-weight ± 2 spherical harmonics (Appendix D). Using this relationship between the tensor spherical harmonic and $(+, \times)$ polarization bases, one can show [69]:

$$h_+(f, \hat{n}) \pm i h_\times(f, \hat{n}) = \frac{1}{\sqrt{2}} \sum_{(lm)} \left(a^G_{(lm)}(f) \pm i a^C_{(lm)}(f) \right) {}_{\pm 2}Y_{lm}(\hat{n}) , \tag{2.10}$$

or, equivalently,

$$a^G_{(lm)}(f) \pm i a^C_{(lm)}(f) = \sqrt{2}\int \mathrm{d}^2\Omega_{\hat{n}}\ \left(h_+(f,\hat{n}) \pm i h_\times(f,\hat{n})\right)\ {}_{\pm 2}Y^*_{lm}(\hat{n})\,. \tag{2.11}$$

These two expressions allow us to go back and forth between the expansion coefficients for the two different bases.

2.3 Statistical properties

The statistical properties of a stochastic gravitational-wave background are specified in terms of the probability distributions of the expansion coefficients, either $h_A(f,\hat{n})$, where $A = \{+,\times\}$ labels the standard polarization modes of general relativity, or $a^P_{(lm)}(f)$, where $P = \{G, C\}$ and (lm) labels the multipole components for the gradient and curl spherical harmonic decomposition. Without loss of generality we can assume that the expansion coefficients have zero mean—i.e.,

$$\langle h_A(f,\hat{n})\rangle = 0\,, \qquad \langle a^P_{(lm)}(f)\rangle = 0\,. \tag{2.12}$$

For Gaussian backgrounds we need only consider quadratic expectation values, since all higher-order moments are either zero or can be written in terms of the quadractic moments (Appendix B). For non-Gaussian backgrounds (Section 8.1), third and/or higher order moments will also be needed.

The specific form of these expectation values will depend on the source of the background. For example, a cosmological background associated with gravitational waves produced during or shortly after inflation would most-likely be described a *stationary* random process, having been produced on time-scales much longer than the duration of an observation (typically of order a few years to ~ 10 years). This cosmological background is also expected to be Gaussian-distributed, being the superposition of many independent gravitational-wave signals (via the central limit theorem), as well as isotropically-distributed on the sky. Contrast this with the superposition of gravitational waves produced by unresolved Galactic white-dwarf binaries radiating in the eLISA band (10^{-4} Hz$-$ 10^{-1} Hz). Although this confusion-limited astrophysical foreground is also expected to be Gaussian and stationary, it will have an *anisotropic distribution*, following the spatial distribution of the Milky Way. The anistropy will be encoded as a modulation in the eLISA output, due to the changing antenna pattern of the eLISA constellation in its yearly orbit around the Sun. Hence, different sources will give rise to different statistical distributions, which we will need to consider when formulating our data analysis strategies.

2.3.1 Quadratic expectation values for Gaussian-stationary backgrounds

The simplest type of stochastic background will be Gaussian-stationary, unpolarized, and spatially homogenous and isotropic. The quadratic expectation values for such a background are then

$$\langle h_A(f,\hat{n}) h^*_{A'}(f',\hat{n}')\rangle = \frac{1}{16\pi} S_h(f)\delta(f-f')\delta_{AA'}\delta^2(\hat{n},\hat{n}')\,, \tag{2.13}$$

or, equivalently,

$$\langle a^{P}_{(lm)}(f) a^{P'*}_{(l'm')}(f') \rangle = \frac{1}{8\pi} S_h(f) \delta(f-f') \delta^{PP'} \delta_{ll'} \delta_{mm'} \,. \tag{2.14}$$

The numerical factors out front have been included so that $S_h(f)$ has the interpretation of being the one-sided gravitational-wave *strain power spectral density* function (units of strain^2/Hz), summed over both polarizations and integrated over the sky. The factor of $\delta(f-f')$ arises due to our assumption of stationarity; the factor of $\delta_{AA'}$ (or $\delta^{PP'}$) is due to our assumption that the polarization modes are statistically independent of one another and have no preferred component; and the factor of $\delta^2(\hat{n}, \hat{n}')$ (or $\delta_{ll'}\delta_{mm'}$) is due to our assumption of spatial homogeneity and isotropy.

Anisotropic, unpolarized, Gaussian-stationary backgrounds, whose radiation from different directions on the sky are uncorrelated with one another, are also simply represented in terms of the quadratic expectation values:

$$\langle h_A(f, \hat{n}) h^*_{A'}(f', \hat{n}') \rangle = \frac{1}{4} \mathcal{P}(f, \hat{n}) \delta(f-f') \delta_{AA'} \delta^2(\hat{n}, \hat{n}') \,. \tag{2.15}$$

The function $\mathcal{P}(f, \hat{n})$ describes the spatial distribution of gravitational-wave power on the sky at frequecy f. It is related to $S_h(f)$ via

$$S_h(f) = \int d^2\Omega_{\hat{n}} \, \mathcal{P}(f, \hat{n}) \,. \tag{2.16}$$

The corresponding expectation values in terms of the tensor spherical harmonic expansion coefficients $a^P_{(lm)}(f)$ are more complicated, since an individual mode in this basis corresponds to a gravitational-wave background whose radiation is correlated between different angular directions on the sky. (See [69] for a discussion of backgrounds that have such correlations.) We will discuss searches for anisotropic backgrounds in more detail in Section 7.

More general Gaussian-stationary backgrounds (e.g., polarized, statistically isotropic but with correlated radiation, etc.) can be represented by appropriately changing the right-hand-side of the quadratic expectation values. However, for the remainder of this section and for most of the article, we will consider "vanilla" isotropic backgrounds, whose quadratic expectation values (2.13) or (2.14) are completely specified by the power spectral density $S_h(f)$.

2.4 Fractional energy density spectrum

The gravitational-wave strain power spectral density $S_h(f)$ is simply related to the fractional energy density spectrum in gravitational waves $\Omega_{\rm gw}(f)$, see e.g., [26]:

$$S_h(f) = \frac{3H_0^2}{2\pi^2} \frac{\Omega_{\rm gw}(f)}{f^3} \,, \tag{2.17}$$

where

$$\Omega_{\rm gw}(f) = \frac{1}{\rho_c} \frac{d\rho_{\rm gw}}{d \ln f} \,. \tag{2.18}$$

Here $d\rho_{\rm gw}$ is the energy density in gravitational waves contained in the frequency interval f to $f+df$, and $\rho_c \equiv 3c^2H_0^2/8\pi G$ is the critical energy density need to close the universe. The *total* energy density in gravitational waves normalised by the critical energy density is thus

$$\Omega_{\rm gw} = \int_{f=0}^{\infty} d(\ln f)\, \Omega_{\rm gw}(f)\,. \tag{2.19}$$

This can be compared, for example, to the total fractional energy density $\Omega_{\rm b}$, Ω_Λ, in baryons, dark energy, etc. Since $\rho_{\rm c}$ involves the Hubble constant, one sometimes writes $H_0 = h_0\, 100~{\rm km~s^{-1}~Mpc^{-1}}$, and then absorbs a factor of h_0^2 in $\Omega_{\rm gw}(f)$. The quantity $h_0^2\,\Omega_{\rm gw}(f)$ is then *independent* of the value of the Hubble constant. However, since recent measurements by Planck [135, 187] have shown that $h_0 = 0.68$ to a high degree of precision, we have assumed this value in this review article and quote limits directly on $\Omega_{\rm gw}(f)$ (Section 10). The specific functional form for $\Omega_{\rm gw}(f)$ depends on the source of the background, as we shall see explicitly below.

2.5 Characteristic strain

Although the fractional energy density spectrum $\Omega_{\rm gw}(f)$ completely characterizes the statistical properties of a Gaussian-stationary isotropic background, it is often convenient to work with the (dimensionless) characteristic strain amplitude $h_c(f)$ defined by

$$h_c(f) \equiv \sqrt{f S_h(f)}\,. \tag{2.20}$$

It is related to $\Omega_{\rm gw}(f)$ via:

$$\Omega_{\rm gw}(f) = \frac{2\pi^2}{3H_0^2} f^2 h_c^2(f)\,. \tag{2.21}$$

Several theoretical models of gravitational-wave backgrounds predict characteristic strains that have a power-law form

$$h_c(f) = A_\alpha \left(\frac{f}{f_{\rm ref}}\right)^\alpha\,, \tag{2.22}$$

where α is spectral index and $f_{\rm ref}$ is typically set to 1/yr. (There is no sum over α in the above expression, and no sum over β in the following expression.) Using equations (2.21) and (2.22) it follows that

$$\Omega_{\rm gw}(f) = \Omega_\beta \left(\frac{f}{f_{\rm ref}}\right)^\beta\,, \tag{2.23}$$

where

$$\Omega_\beta = \frac{2\pi^2}{3H_0^2} f_{\rm ref}^2\, A_\alpha^2\,, \quad \beta = 2\alpha + 2\,. \tag{2.24}$$

For inflationary backgrounds relevant for cosmology, it is often assumed that $\Omega_{\rm gw}(f) =$ const, for which $\beta = 0$ and $\alpha = -1$. For a background arising from binary coalescence, $\Omega_{\rm gw}(f) \propto f^{2/3}$, for which $\beta = 2/3$ and $\alpha = -2/3$. This power-law dependence is applicable to super-massive black-hole binary (SMBHB) coalescences targeted by pulsar timing observations as well as to compact binary coalescences relevant for ground-based and space-based detectors.

3 Statistical inference

> If your experiment needs statistics, you ought to have done a better experiment. *Ernest Rutherford*

In this section, we review statistical inference from both the Bayesian and frequentist perspectives. Our discussion of frequentist and Bayesian upper limits, and the example given in Section 3.5 comparing Bayesian and frequentist analyses is modelled in part after [146]. Readers interested in more details about Bayesian statistical inference should see e.g., [88, 89, 96, 77, 160]. For a description of frequentist statistics, we recommend [83, 195, 64]. Readers who already familiar with the topic of statistical inference can skip this section and go directly to Section 4, although it might still be worthwhile to skim through it, just to see the notation that we will be using in the rest of the article.

3.1 Introduction to Bayesian and frequentist inference

Statistical inference can be used to answer questions such as "Is a gravitational-wave signal present in the data?" and, if so, "What are the physical characteristics of the source?" These questions are addressed using the techniques of classical (also known as *frequentist*) inference and *Bayesian* inference. Many of the early theoretical studies and observational papers in gravitational-wave astronomy followed the frequentist approach, but the use of Bayesian inference is growing in popularity. Moreover, many contemporary analyses cannot be classified as purely frequentist or Bayesian.

The textbook definition states that the difference between the two approaches comes down to their different interpretations of probability: for frequentists, probabilities are fundamentally related to frequencies of events, while for Bayesians, probabilities are fundamentally related to our own knowledge about an event. For example, when inferring the mass of a star, the frequentist interpretation is that the star has a true, fixed (albeit unknown) mass, so it is meaningless to talk about a probability distribution for it. Rather, the uncertainty is in the data, and the relevant probability is that of observing the data d, given that the star has mass m. This probability distribution is the *likelihood*, denoted $p(d|m)$. In contrast, in the Bayesian interpretation the data are known (after all, it is what is measured!), and the mass of the star is what we are uncertain about[3], so the relevant probability is that the mass has a certain value, given the data. This probability distribution is the *posterior*, $p(m|d)$. The likelihood and posterior are related via Bayes' theorem:

$$p(m|d) = \frac{p(d|m)p(m)}{p(d)}\,, \tag{3.1}$$

[3] In some treatments the Bayesian interpretation is equated to philosophical schools such as Berkeley's empiricist idealism, or subjectivism, which holds things only exist to the extent that they are perceived, while the frequentist interpretation is equated to Platonic realism, or metaphysical objectivism, holding that things exist objectively and independently of observation. These equivalences are false. A physical object can have a definite, Platonic existence, and Bayesians can still assign probabilities to its attributes since our ability to measure is limited by imperfect equipment.

where $p(m)$ is the prior probability distribution for m, and the normalization constant,

$$p(d) = \int p(d|m)p(m)\, dm \,, \tag{3.2}$$

is the *marginalized likelihood*, or *evidence*. For uniform (flat) priors the frequentist confidence intervals for the parameters will coincide with the Bayesian credible intervals, but the interpretation remains quiet distinct.

The choice of prior probability distributions is a source of much consternation and debate, and is often cited as a weakness of the Bayesian approach. But the choice of probability distribution for the likelihood (which is also important for the frequentist approach) is often no less fraught. The prior quantifies what we know about the range and distribution of the parameters in our model, while the likelihood quantifies what we know about our measurement apparatus, and, in particular, the nature of the measurement noise. The choice of prior is especially problematic in a new field where there is little to guide the choice. For example, electromagnetic observations and population synthesis models give some guidance about black hole masses, but the mass range and distribution is currently not well constrained. The choice of likelihood can also be challenging when the measurement noise deviates from the stationary, Gaussian ideal. More details related to the choice of likelihood and choice of prior will be given in Section 3.6.

In addition to parameter estimation, statistical inference is used to select between competing models, or hypotheses, such as, "is there a gravitational-wave signal in the data or not?" Thanks to GW150914 and GW151226, we know that gravitational-wave signals *are* already present in existing data sets, but most are at levels where we are unable to distinguish them from noise processes. For detection we demand that a model for the data that includes a gravitational-wave signal be favored over a model having no gravitational-wave signal. In Bayesian inference a detection might be announced when the odds ratio between models with and without gravitational-wave signals gets sufficiently large, while in frequentist inference a detection might be announced when the p-value for some test statistic is less than some prescribed threshold. These different approaches to deciding whether or not to claim a detection (e.g., Bayesian model selection or frequentist hypothesis testing), as well as differences in regard to parameter estimation, are described in the following subsections. Table 2 provides an overview of the key similarities and differences between frequentist and Bayesian inference, to be described in detail below.

3.2 Frequentist statistics

As mentioned above, classical or *frequentist* statistics is a branch of statistical inference that interprets probability as the "long-run relative occurrence of an event in a set of identical experiments." Thus, for a frequentist, probabilities can only be assigned to propositions about outcomes of (in principle) repeated experiments (i.e., *random variables*) and not to hypotheses or parameters describing the state of nature, which have fixed but unknown values. In this interpretation, the measured data are drawn from an underlying probability distribution, which assumes the truth of a particular hypothesis or model. The probability distribution for the data is just the likelihood function, which we can write as $p(d|H)$, where d denotes the data and H denotes an hypothesis.

FREQUENTIST	BAYESIAN
probabilities assigned only to propositions about outcomes of repeatable experiments (i.e., random variables), not to hypotheses or parameters which have fixed but unknown values	probabilities can be assigned to hypotheses and parameters since probability is degree of belief (or confidence, plausibility) in any proposition
assumes measured data are drawn from an underlying probability distribution, which assumes the truth of a particular hypothesis or model (likelihood function)	same
constructs a statistic to estimate a parameter or to decide whether or not to claim a detection	needs to specify prior degree of belief in a particular hypothesis or parameter
calculates the probability distribution of the statistic (sampling distribution)	uses Bayes' theorem to update the prior degree of belief in light of new data (i.e., likelihood "plus" prior yields posterior)
constructs confidence intervals and p-values for parameter estimation and hypothesis testing	constructs posteriors and odds ratios for parameter estimation and hypothesis testing / model comparison

Table 2: Comparison of frequentist and Bayesian approaches to statistical inference. See Sections 3.2 and 3.3 for details.

Statistics play an important role in the frequentist framework. These are random variables constructed from the data, which typically estimate a signal parameter or indicate how well the data fits a particular hypothesis. Although it is common to construct statistics from the likelihood function (e.g., the maximum-likelihood statistic for a particular parameter, or the maximum-likelihood ratio to compare a signal-plus-noise model to a noise-only model), there is no a priori restriction on the form of a statistic other than it be *some* function of the data. Ultimately, it is the goal of the analysis and the cleverness of the data analyst that dictate which statistic (or statistics) to use.

To make statistical inferences in the frequentist framework requires knowledge of the probability distribution (also called the *sampling distribution*) of the statistic. The sampling distribution can either be calculated analytically (if the statistic is sufficiently simple) or via Monte Carlo simulations, which effectively constructs a histogram of the values of the statistic by simulating many independent realizations of the data. Given a statistic and its sampling distribution, one can then calculate either *confidence intervals* for parameter estimation or *p-values* for hypothesis testing. (These will be discussed in more detail below.) Note that a potential problem with frequentist statistical inference is that the sampling distribution depends on data values that were *not* actually observed, which is related to how the experiment was carried out *or might have been* carried out. The so-called *stopping problem* of frequentist statistics is an example of such a problem [89].

3.2.1 Frequentist hypothesis testing

Suppose, as a frequentist, you want to test the hypothesis H_1 that a gravitational-wave signal, having some fixed but unknown amplitude $a > 0$, is present in the data. Since you cannot assign probabilities to hypotheses or to parameters like a as a frequentist, you need to introduce instead an alternative (or *null*) hypothesis H_0, which, for this example, is the hypothesis that there is no gravitational-wave signal in the data (i.e., that $a = 0$). You then argue for H_1 by arguing *against* H_0, similar to proof by contradiction in mathematics. Note that H_1 is a *composite* hypothesis since it depends on a range of values of the unknown parameter a. It can be written as the union, $H_1 = \cup_{a>0} H_a$, of a set of simple hypotheses H_a each corresponding to a single fixed value of the parameter a.

To rule either in favor or against H_0, you construct a statistic Λ, called a *test* or *detection statistic*, on which the statistical test will be based. As mentioned above, you will need to calculate analytically or via Monte Carlo simulations the sampling distribution for Λ under the assumption that the null hypothesis is true, $p(\Lambda|H_0)$. If the observed value of Λ lies far out in the tails of the distribution, then the data are most likely not consistent with the assumption of the null hypothesis, so you reject H_0 (and thus accept H_1) at the $p * 100\%$ level, where

$$p \equiv \mathrm{Prob}(\Lambda > \Lambda_{\mathrm{obs}}|H_0) \equiv \int_{\Lambda_{\mathrm{obs}}}^{\infty} p(\Lambda|H_0)\, d\Lambda\,. \tag{3.3}$$

This is the so-called *p-value* (or *significance*) of the test; it is illustrated graphically in Figure 3. The p-value required to reject the null hypothesis determines a *threshold* Λ_*, above which you reject H_0 and accept H_1 (e.g., claim a detection). It is related to the *false alarm probability* for the test as we explain below.

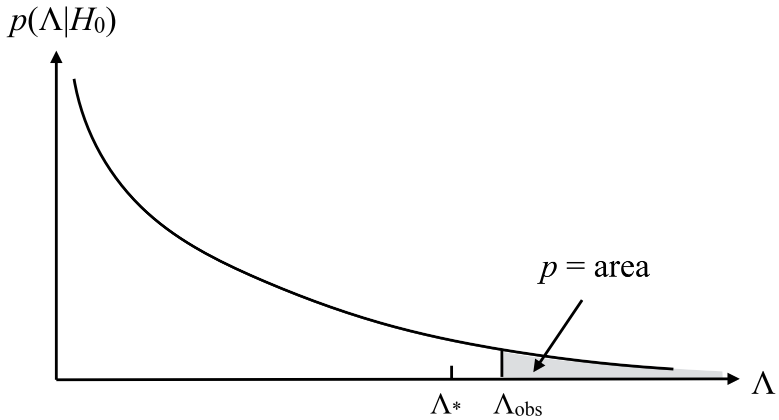


Figure 3: Definition of the p-value (or significance) for frequentist hypothesis testing. The value of p equals the area under the probability distribution $p(\Lambda|H_0)$ for $\Lambda \geq \Lambda_{\rm obs}$.

The above statistical test is subject to two types of errors: (i) type I or *false alarm* errors, which arise if the data are such that you reject the null hypothesis (i.e., $\Lambda_{\rm obs} > \Lambda_*$) when it is actually true, and (ii) type II or *false dismissal* errors, which arise if the data are such that you accept the null hypothesis (i.e., $\Lambda_{\rm obs} < \Lambda_*$) when it is actually false. The false alarm probability α and false dismissal probability $\beta(a)$ are given explicitly by

$$\alpha \equiv {\rm Prob}(\Lambda > \Lambda_*|H_0)\,, \tag{3.4}$$

$$\beta(a) \equiv {\rm Prob}(\Lambda < \Lambda_*|H_a)\,, \tag{3.5}$$

where a is the amplitude of the gravitational-wave signal, assumed to be present under the assumption that H_1 is true. To calculate the false dismissal probability $\beta(a)$, one needs the sampling distribution of the test statistic assuming the presence of a signal with amplitude a.

Different test statistics are judged according to their false alarm and false dismissal probabilities. Ideally, you would like your statistical test to have false alarm and false dismissal probabilities that are both as small as possible. But these two properties compete with one another as setting a larger threshold value to minimize the false alarm probability will increase the false dismissal probability. Conversely, setting a smaller threshold value to minimize the false dismissal probability will increase the false alarm probability.

In the context of gravitational-wave data analysis, the gravitational-wave community is (at least initially) reluctant to falsely claim detections. Hence the false alarm probability is set to some very low value. The best statistic then is the one that minimizes the false dismissal probability (i.e., maximizes detection probability) for fixed false alarm. This is the *Neyman-Pearson criterion*. For medical diagnosis, on the other hand, a doctor is very reluctant to falsely dismiss an illness. Hence the false dismissal probability will be set to some very low value. The best statistic then is the one which minimizes the false alarm probability for fixed false dismissal.

3.2.2 Frequentist detection probability

The value $1-\beta(a)$ is called the *detection probability* or *power* of the test. It is the fraction of times that the test stastic Λ correctly identifies the presence of a signal of amplitude a in the data, for a fixed false alarm probability α (which sets the threshold Λ_*). A plot of detection probability versus signal strength is often used to show how strong a signal has to be in order to detect it with a certain probability. Since detection probability does not depend on the observed data—it depends only on the sampling distributions of the test statistic and a choice for the false alarm probability—detection probability curves are often used as a figure-of-merit for proposed search methods for a signal. Figure 4 shows a detection probability curve, with the value of a needed to be detectable with 90% frequentist probability indicated by the dashed vertical line. We will denote this value of a by $a^{90\%,\mathrm{DP}}$. Note that as the signal amplitude goes to zero, the detection probability reduces to the false alarm probability α, which for this example was chosen to be 0.10.

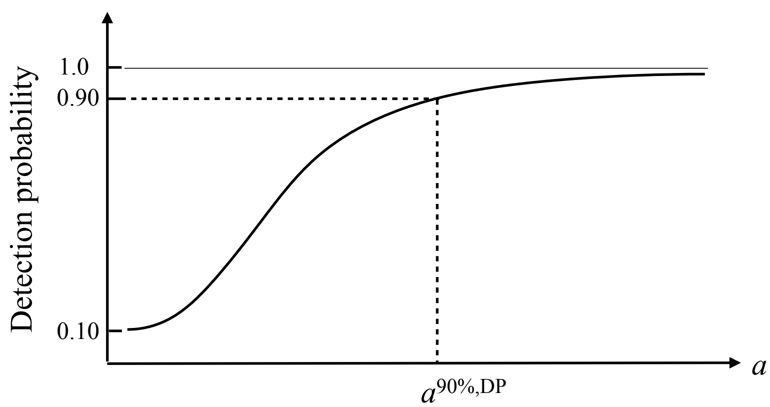


Figure 4: Detection probability as a function of the signal amplitude for a false alarm probability equal to 10%. The value of a needed for 90% detection probability is indicated by the dashed vertical line and is denoted by $a^{90\%,\mathrm{DP}}$.

3.2.3 Frequentist upper limits

In the absence of a detection (i.e., if the observed value of the test statistic is less than the detection threshold Λ_*), one can still set a bound (called an *upper limit*) on the strength of the signal that one was trying to detect. The upper limit depends on the observed value of the test statistic, Λ_{obs}, and a choice of confidence level, CL, interpreted in the frequentist framework as the long-run relative occurence for a set of repeated identical experiments. For example, one defines the 90% confidence-level upper limit $a^{90\%,\mathrm{UL}}$ as the minimum value of a for which $\Lambda \geq \Lambda_{\mathrm{obs}}$ at least 90% of the time:

$$\mathrm{Prob}(\Lambda \geq \Lambda_{\mathrm{obs}} | a \geq a^{90\%,\mathrm{UL}}, H_a) \geq 0.90\,. \tag{3.6}$$

In other words, if the signal has an amplitude $a^{90\%,\mathrm{UL}}$ or higher, we would have detected it in 90% of repeated observations. A graphical representation of a frequentist upper limit is given in Figure 5.

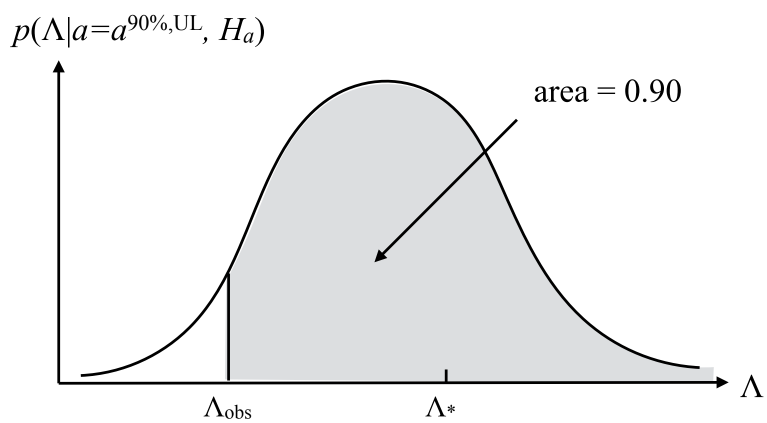


Figure 5: Graphical representation of a frequentist 90% confidence level upper limit. When $a = a^{90\%,\mathrm{UL}}$, the probability of obtaining a value of the detection statistic $\Lambda \geq \Lambda_{\mathrm{obs}}$ is equal to 0.90.

3.2.4 Frequentist parameter estimation

The frequentist prescription for estimating the value of a particular parameter a, like the amplitude of a gravitational-wave signal, is slightly different than the method used to claim a detection. You need to first construct a statistic (called an *estimator*) $\hat{a}$ of the parameter a you are interested in. (This might be a maximum-likelihood estimator of a, but other estimators can also be used.) You then calculate its sampling distribution $p(\hat{a}|a, H_a)$. Note that statements like

$$\mathrm{Prob}(a - \Delta < \hat{a} < a + \Delta) = 0.95\,, \tag{3.7}$$

which one constructs from $p(\hat{a}|a, H_a)$ make sense in the frequentist framework, since $\hat{a}$ is a random variable. Although the above inequality can be rearranged to yield

$$\mathrm{Prob}(\hat{a} - \Delta < a < \hat{a} + \Delta) = 0.95\,, \tag{3.8}$$

this should *not* be interpreted as a statement about the probability of a lying within a particular interval $[\hat{a} - \Delta, \hat{a} + \Delta]$, since a is not a random variable. Rather, it should be interpreted as a probabilistic statement about the *set of intervals* $\{[\hat{a} - \Delta, \hat{a} + \Delta]\}$ for all possible values of $\hat{a}$. Namely, in a set of many repeated experiments, 0.95 is the fraction of the intervals that will contain the true value of the parameter a. Such an interval is called a 95% *frequentist confidence interval*. This is illustrated graphically in Figure 6.

It is important to point out that an estimator can sometimes take on a value of the parameter that is *not physically allowed*. For example, if the parameter a denotes the

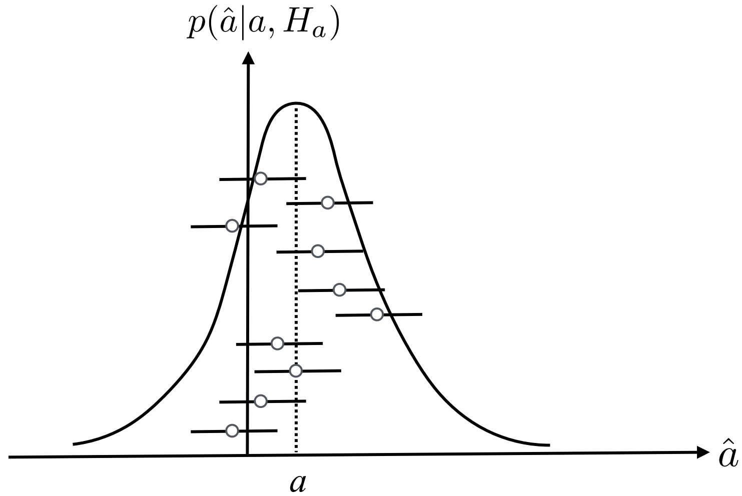


Figure 6: Definition of the frequentist confidence interval for parameter estimation. Each circle and line represents a measured interval $[\hat{a} - \Delta, \hat{a} + \Delta]$. The set of all such intervals will contain the true value of the parameter a (indicated here by the dotted vertical line) $\mathrm{CL} * 100\%$ of the time, where CL is the confidence level.

amplitude of a gravitational-wave signal (so physically $a \geq 0$), it is possible for $\hat{a} < 0$ for a particular realization of the data. Note that there is nothing mathematically wrong with this result. Indeed, the sampling distribution for $\hat{a}$ specifies the probability of obtaining such values of $\hat{a}$. It is even possible to have a confidence interval $[\hat{a}-\Delta, \hat{a}+\Delta]$ all of whose values are unphysical, especially if one is trying to detect a weak signal in noise. Again, this is mathematically allowed, but it is a little awkward to report a frequentist confidence interval that is completely unphysical. We shall see that within the Bayesian framework unphysical intervals and unphysical posteriors never arise, as a simple consequence of including a prior distribution on the parameter that requires $a \geq 0$.

3.3 Bayesian inference

In the following subsections, we again describe parameter estimation and hypothesis testing, but this time from the perspective of Bayesian inference.

3.3.1 Bayesian parameter estimation

In Bayesian inference, a parameter, e.g., a, is estimated in terms of its posterior distribution, $p(a|d)$, in light of the observed data d. As discussed in the introduction to this section, the posterior $p(a|d)$ can be calculated from the likelihood $p(d|a)$ and the prior probability distribution $p(a)$ using Bayes' theorem

$$p(a|d) = \frac{p(d|a)p(a)}{p(d)} \,. \tag{3.9}$$

The posterior distribution tells you everything you need to know about the parameter, although you might sometimes want to reduce it to a few numbers—e.g., its mode, mean, standard deviation, etc.

Given a posterior distribution $p(a|d)$, a Bayesian confidence interval (often called a *credible interval* given the Bayesian interpretation of probability as degree of belief, or state of knowledge, about an event) is simply defined in terms of the area under the posterior between one parameter value and another. This is illustrated graphically in Figure 7, for the case of a 95% symmetric credible interval, centered on the mode of the distribution $a_{\rm mode}$. If the posterior distribution depends on two parameters a and b, but

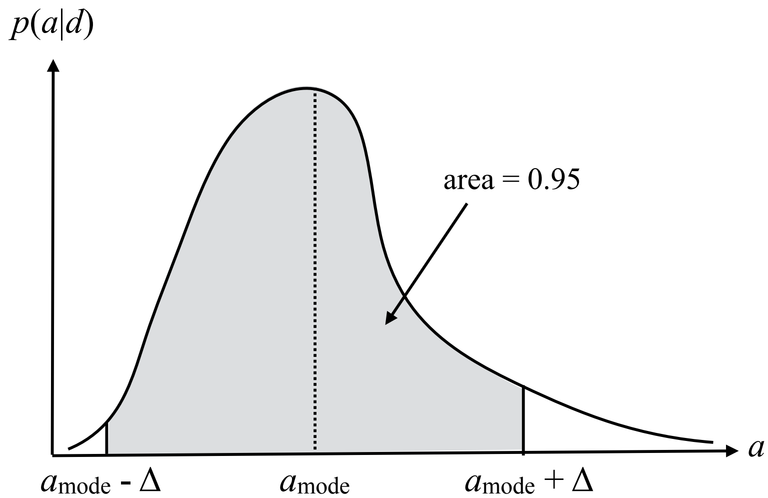


Figure 7: Definition of a Bayesian credible interval for parameter estimation. Here we construct a symmetric 95% credible interval centered on the mode of the distribution.

you really only care about a, then you can obtain the posterior distribution for a by marginalising the joint distribution $p(a, b|d)$ over b:

$$p(a|d) = \int db\, p(a, b|d) = \int db\, p(a|b, d)p(b)\,, \tag{3.10}$$

where the second equality follows from the relationship between joint probabilities and conditional probabilities, e.g., $p(a|b, d)p(b) = p(a, b|d)$. Variables that you don't particularly care about (e.g., the variance of the detector noise as opposed to the strength of a gravitational-wave signal) are called *nuisance parameters*. Although nuisance parameters can be handled in a straight-forward manner using Bayesian inference, they are problematic to deal with (i.e., they are a nuisance!) in the context of frequentist statistics. The problem is that marginalization doesn't make sense to a frequentist, for whom parameters cannot be assigned probability distributions.

The interpretation of Bayes' theorem (3.9) is that our prior knowledge is updated by what we learn from the data, as measured by the likelihood, to give our posterior state of knowledge. The amount learned from the data is measured by the information gain

$$I = \int da\, p(a|d) \log \left(\frac{p(a|d)}{p(a)} \right) . \tag{3.11}$$

Using a natural logarithm gives the information in *nats*, while using a base 2 logarithm gives the information in *bits*. If the data tells us nothing about the parameter, then $p(d|a) = \text{constant}$, which implies $p(a|d) = p(a)$ and thus $I = 0$.

3.3.2 Bayesian upper limits

A Bayesian upper limit is simply a Bayesian credible interval for a parameter with the lower end point of the interval set to the smallest value that the parameter can take. For example, the Bayesian 90% upper limit on a parameter $a > 0$ is defined by:

$$\text{Prob}(0 < a < a^{90\%,\text{UL}}|d) = 0.90\,, \tag{3.12}$$

where probability is interpreted as degree of belief, or state of knowledge, that the parameter a has a value in the indicated range. One usually sets an upper limit on a parameter when the mode of the distribution for the parameter being estimated is not sufficiently displaced from zero, as shown in Figure 8.

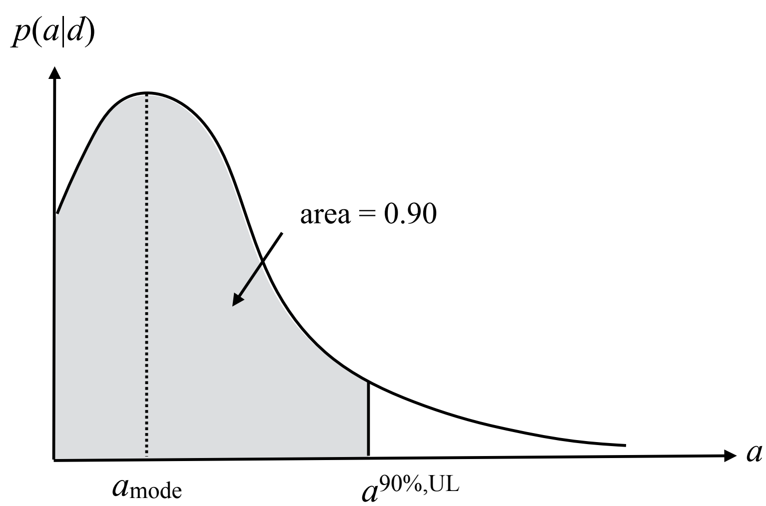


Figure 8: Bayesian 90% credible upper limit for the parameter a.

3.3.3 Bayesian model selection

Bayesian inference can easily be applied to multiple models or hypotheses, each with a different set of parameters. In what follows, we will denote the different models by $\mathcal{M}_\alpha$, where the index α runs over the different models, and the associated set of parameters by the vector $\vec{\theta}_\alpha$. The joint posterior distribution for the parameters $\vec{\theta}_\alpha$ is given by

$$p(\vec{\theta}_\alpha|d, \mathcal{M}_\alpha) = \frac{p(d|\vec{\theta}_\alpha, \mathcal{M}_\alpha)p(\vec{\theta}_\alpha|\mathcal{M}_\alpha)}{p(d|\mathcal{M}_\alpha)}\,, \tag{3.13}$$

and the model evidence is given by

$$p(d|\mathcal{M}_\alpha) = \int p(d|\vec{\theta}_\alpha, \mathcal{M}_\alpha)p(\vec{\theta}_\alpha|\mathcal{M}_\alpha)\, d\vec{\theta}_\alpha\,, \tag{3.14}$$

where we marginalize over the parameter values associated with that model. The posterior probability for model $\mathcal{M}_\alpha$ is given by Bayes' theorem as

$$p(\mathcal{M}_\alpha|d) = \frac{p(d|\mathcal{M}_\alpha)p(\mathcal{M}_\alpha)}{p(d)}\,, \tag{3.15}$$

where the normalization constant $p(d)$ involves a sum over all possible models:

$$p(d) = \sum_\alpha p(d|\mathcal{M}_\alpha)p(\mathcal{M}_\alpha)\,. \tag{3.16}$$

Since the space of all possible models is generally unknown, the sum is usually taken over the subset of models being considered. The normalization can be avoided by considering the posterior odds ratio between two models:

$$\mathcal{O}_{\alpha\beta}(d) = \frac{p(\mathcal{M}_\alpha|d)}{p(\mathcal{M}_\beta|d)} = \frac{p(\mathcal{M}_\alpha)}{p(\mathcal{M}_\beta)}\,\frac{p(d|\mathcal{M}_\alpha)}{p(d|\mathcal{M}_\beta)}\,. \tag{3.17}$$

The first ratio on the right-hand side of the above equation is the *prior* odds ratio for models α, β, while the second term is the evidence ratio, or *Bayes factor*,

$$\mathcal{B}_{\alpha\beta}(d) \equiv \frac{p(d|\mathcal{M}_\alpha)}{p(d|\mathcal{M}_\beta)}\,. \tag{3.18}$$

The prior odds ratio is often taken to equal unity, but this is not always justified. For example, the prior odds that a signal is described by General Relativity versus some alternative theory of gravity should be much larger than unity given the firm theoretical and observational footing of Einstein's theory.

While the foundations of Bayesian inference were laid out by Laplace in the 1700s, it did not see widespread use until the late 20th century with the advent of practical implementation schemes and the development of fast electronic computers. Today, Monte Carlo sampling techniques, such as Markov Chain Monte Carlo (MCMC) and Nested Sampling, are used to sample the posterior and estimate the evidence [161, 70]. Successfully applying these techniques is something of an art, but in principle, once the likelihood and prior have been written down, the implementation of Bayesian inference is purely mechanical. Calculating the likelihood and choosing a prior will be discussed in some detail in Section 3.6.

3.4 Relating Bayesian and frequentist detection statements

It is interesting to compare the Bayesian model selection calculation discussed above to frequentist hypothesis testing based on the *maximum-likelihood ratio*. For concreteness, let's assume that we have two models $\mathcal{M}_0$ (noise-only) and $\mathcal{M}_1$ (noise plus gravitational-wave signal), with parameters $\vec{\theta}_n$ and $\{\vec{\theta}_n, \vec{\theta}_h\}$, respectively. The frequentist detection statistic will be defined in terms of the ratio of the maxima of the likelihood functions for the two models:

$$\Lambda_{\rm ML}(d) \equiv \frac{\max_{\vec{\theta}_n} \max_{\vec{\theta}_h} p(d|\vec{\theta}_n, \vec{\theta}_h, \mathcal{M}_1)}{\max_{\vec{\theta}'_n} p(d|\vec{\theta}'_n, \mathcal{M}_0)} \,. \tag{3.19}$$

As described above, the Bayes factor calculation also involves a ratio of two quantities, the model evidences $p(d|\mathcal{M}_1)$ and $p(d|\mathcal{M}_0)$, but instead of maximizing over the parameters, we marginalize over the parameters:

$$\mathcal{B}_{10}(d) = \frac{\int d\vec{\theta}_n \int d\vec{\theta}_h \, p(d|\vec{\theta}_n, \vec{\theta}_h, \mathcal{M}_1) p(\vec{\theta}_n, \vec{\theta}_h|\mathcal{M}_1)}{\int d\vec{\theta}'_n \, p(d|\vec{\theta}'_n, \mathcal{M}_0) p(\vec{\theta}'_n|\mathcal{M}_0)} \,. \tag{3.20}$$

These two expressions can be related using Laplace's approximation to individually approximate the model evidences $p(d|\mathcal{M}_1)$ and $p(d|\mathcal{M}_0)$. This approximation is valid when the data are *informative*—i.e., when the likelihood functions are peaked relative to the joint prior probability distributions of the parameters. For an arbitrary model $\mathcal{M}$ with parameters $\vec{\theta}$, the Laplace approximation yields:

$$\int d\vec{\theta} \, p(d|\vec{\theta}, \mathcal{M}) \simeq p(d|\vec{\theta}_{\rm ML}, \mathcal{M}) \frac{\Delta V_{\mathcal{M}}}{V_{\mathcal{M}}} \,, \tag{3.21}$$

where $\vec{\theta}_{\rm ML} \equiv \vec{\theta}_{\rm ML}(d)$ maximizes the likelihood with respect to variation of $\vec{\theta}$ given the data d; $\Delta V_{\mathcal{M}}$ is the characteristic width of the likelihood function around its maximum (the volume of the uncertainty ellipsoid for the parameters); and $V_{\mathcal{M}}$ is the total parameter space volume of the model parameters. Applying this approximation to models $\mathcal{M}_0$ and $\mathcal{M}_1$ in (3.20), we obtain

$$\mathcal{B}_{10}(d) \simeq \Lambda_{\rm ML}(d) \frac{\Delta V_1/V_1}{\Delta V_0/V_0} \,, \tag{3.22}$$

or, equivalently,

$$2 \ln \mathcal{B}_{10}(d) \simeq 2 \ln \left(\Lambda_{\rm ML}(d)\right) + 2 \ln \left(\frac{\Delta V_1/V_1}{\Delta V_0/V_0}\right) \,. \tag{3.23}$$

The second term in the above equation is negative and penalizes models that require a larger parameter space volume than necessary to fit the data. This is basically an *Occam penalty factor*, which prefers the simpler of two models that fit the data equally well. The first term has the interpretation of being the squared signal-to-noise ratio of the data, assuming an additive signal in Gaussian-stationary noise, and it can be used as an alternative frequentist detection statistic in place of $\Lambda_{\rm ML}$.

Table 3 from [100] gives a range of Bayes factors and their interpretation in terms of the strength of the evidence in favor of one model relative to another. The precise levels

$\mathcal{B}_{\alpha\beta}(d)$	$2\ln\mathcal{B}_{\alpha\beta}(d)$	Evidence for model $\mathcal{M}_\alpha$ relative to $\mathcal{M}_\beta$
< 1	< 0	Negative (supports model $\mathcal{M}_\beta$)
1–3	0–2	Not worth more than a bare mention
3–20	2–6	Positive
20–150	6–10	Strong
> 150	> 10	Very strong

Table 3: Bayes factors and their interpretation in terms of the strength of the evidence in favor of one model relative to the other. Adapted from [100].

at which one considers the evidence to be "strong" or "very strong" is rather subjective. But recent studies [50, 167] in the context of pulsar timing have been trying to make this correspondence a bit firmer, using *sky* and *phase scrambles* to effectively destroy signal-induced spatial correlations between pulsars while retaining the statistical properties of each individual dataset. This is similar to doing time-slides for LIGO analyses, which are used to assess the significance of a detection.

3.5 Simple example comparing Bayesian and frequentist analyses

To further illustrate the relationship between Bayesian and frequentist analyses, we consider in this section a very simple example—a constant signal with amplitude $a > 0$ in white, Gaussian noise (zero mean, variance σ):

$$d_i = a + n_i\,, \qquad i = 1, 2, \cdots, N\,, \tag{3.24}$$

where the index i labels the individual samples of the data. The likelihood functions for the noise-only and signal-plus-noise models $\mathcal{M}_0$ and $\mathcal{M}_1$ are thus simple Gaussians:

$$p(d|\mathcal{M}_0) = \frac{1}{(2\pi)^{N/2}\sigma^N} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^N d_i^2}\,, \tag{3.25}$$

$$p(d|a, \mathcal{M}_1) = \frac{1}{(2\pi)^{N/2}\sigma^N} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^N (d_i-a)^2}\,. \tag{3.26}$$

We will assume that the value of σ is known a priori. Thus, the noise model has no free parameters, while the signal model has just one parameter, which is the amplitude of the signal that we are trying to detect. We will choose our prior on a to be flat over the interval $[0, a_{\rm max}]$, so $p(a) = 1/a_{\rm max}$.

It is straight-forward exercise to check that the maximum-likelihood estimator of the amplitude a is given by the sample mean of the data:

$$\hat{a} \equiv a_{\rm ML}(d) = \frac{1}{N}\sum_{i=1}^N d_i \equiv \bar{d}\,. \tag{3.27}$$

This is is an unbiased estimator of a and has variance $\sigma_{\hat{a}}^2 = \sigma^2/N$ (the familiar variance of the sample mean). Thus, the sampling distribution of $\hat{a}$ is simply:

$$p(\hat{a}|a, \mathcal{M}_1) = \frac{1}{\sqrt{2\pi}\sigma_{\hat{a}}} e^{-\frac{1}{2\sigma_{\hat{a}}^2}(\hat{a}-a)^2}\,, \tag{3.28}$$

where $\hat{a}$ can take on either positive or negative values (even though $a > 0$).

To compute the posterior distribution $p(a|d, \mathcal{M}_1)$ for the Bayesian analysis, we first note that

$$\sum_{i=1}^{N}(d_i - a)^2 = N(\mathrm{Var}[d] + (a - \hat{a})^2)\,. \tag{3.29}$$

The model evidence $p(d|\mathcal{M}_1)$ is then given by

$$p(d|\mathcal{M}_1) = \frac{e^{-\frac{\mathrm{Var}[d]}{2\sigma_{\hat{a}}^2}}\left[\mathrm{erf}\left(\frac{a_{\max}-\hat{a}}{\sqrt{2}\sigma_{\hat{a}}}\right) + \mathrm{erf}\left(\frac{\hat{a}}{\sqrt{2}\sigma_{\hat{a}}}\right)\right]}{2a_{\max}\sqrt{N}(2\pi)^{(N-1)/2}\sigma^{(N-1)}}\,, \tag{3.30}$$

and the posterior distribution is given by

$$p(a|d, \mathcal{M}_1) = \frac{1}{\sqrt{2\pi}\sigma_{\hat{a}}}e^{-\frac{(a-\hat{a})^2}{2\sigma_{\hat{a}}^2}}\,2\left[\mathrm{erf}\left(\frac{a_{\max}-\hat{a}}{\sqrt{2}\sigma_{\hat{a}}}\right) + \mathrm{erf}\left(\frac{\hat{a}}{\sqrt{2}\sigma_{\hat{a}}}\right)\right]^{-1}\,. \tag{3.31}$$

Note that this is simply a *truncated* Gaussian on the interval $a \in [0, a_{\max}]$, with mean $\hat{a}$ and variance $\sigma_{\hat{a}}^2$.

The above calculation shows that $\hat{a}$ is a *sufficient statisitic* for a. This means that the posterior distribution for a can be written simply in terms of $\hat{a}$, in lieu of the individual samples $d \equiv \{d_1, d_2, \cdots, d_N\}$. The Bayes factor

$$\mathcal{B}_{10}(d) = \frac{p(d|\mathcal{M}_1)}{p(d|\mathcal{M}_0)}\,, \tag{3.32}$$

is given by

$$\mathcal{B}_{10}(d) = e^{\frac{\hat{a}^2}{2\sigma_{\hat{a}}^2}}\left(\frac{\sqrt{2\pi}\sigma_{\hat{a}}}{a_{\max}}\right)\frac{1}{2}\left[\mathrm{erf}\left(\frac{a_{\max}-\hat{a}}{\sqrt{2}\sigma_{\hat{a}}}\right) + \mathrm{erf}\left(\frac{\hat{a}}{\sqrt{2}\sigma_{\hat{a}}}\right)\right]\,. \tag{3.33}$$

In the limit where $\hat{a}$ is tightly peaked away from 0 and $a_{\max}$, the Bayes factor simplifies to

$$\mathcal{B}_{10}(d) \simeq e^{\frac{\hat{a}^2}{2\sigma_{\hat{a}}^2}}\left(\frac{\sqrt{2\pi}\sigma_{\hat{a}}}{a_{\max}}\right)\,. \tag{3.34}$$

If we take the frequentist detection statistic to be twice the log of the maximum-likelihood ratio, $\Lambda(d) \equiv 2\ln\Lambda_{\mathrm{ML}}(d)$, then

$$\Lambda(d) = \frac{\hat{a}^2}{\sigma_{\hat{a}}^2} = \frac{\bar{d}^2}{\sigma^2/N} \equiv \rho^2\,, \tag{3.35}$$

which is just the squared signal-to-noise ratio of the data. Furthermore, taking twice the log of the approximate Bayes factor in (3.34) gives

$$2\ln\mathcal{B}_{10}(d) \simeq \Lambda(d) + 2\ln\left(\frac{\sqrt{2\pi}\sigma_{\hat{a}}}{a_{\max}}\right)\,, \tag{3.36}$$

where the first term is just the frequentist detection statistic and second term expresses the Occam penalty. This last result is consistent with the general relation (3.23) discussed in the previous subsection.

The statistical distribution of the frequentist detection statistic can be found in closed form for this simple example. Since a linear combination of Gaussian random variables is also Gaussian-distributed, Λ is the *square* of a (single) Gaussian random variable $\rho = \bar{d}\sqrt{N}/\sigma$. Moreover, since ρ has mean $\mu \equiv a\sqrt{N}/\sigma$ and unit variance, the sampling distribution for Λ in the presence of a signal is a *noncentral chi-squared* distribution with one degree of freedom and non-centrality parameter $\lambda \equiv \mu^2 = a^2 N/\sigma^2$:

$$p(\Lambda|a, \mathcal{M}_1) = \frac{1}{2} e^{-(\Lambda+\lambda)/2} \left(\frac{\Lambda}{\lambda}\right)^{-1/4} I_{-1/2}(\sqrt{\lambda\Lambda})\,, \tag{3.37}$$

where $I_{-1/2}$ is a modified Bessel function of the first kind of order $-1/2$. In the absence of a signal (i.e., when a and hence λ are equal to zero), Λ is given by an (ordinary) chi-squared distribution with one degree of freedom:

$$p(\Lambda|\mathcal{M}_0) = \frac{1}{\sqrt{2}\Gamma(1/2)} \Lambda^{-1/2} e^{-\Lambda/2}\,, \tag{3.38}$$

where Γ is the gamma function. Substituting explicit expressions for $I_{-1/2}(\sqrt{\lambda\Lambda})$ and $\Gamma(1/2)$, we find:

$$p(\Lambda|\mathcal{M}_0) = \frac{1}{\sqrt{2\pi\Lambda}} e^{-\Lambda/2}\,, \tag{3.39}$$

$$p(\Lambda|a, \mathcal{M}_1) = \frac{1}{\sqrt{2\pi\Lambda}} \frac{1}{2} \left[e^{-\frac{1}{2}(\sqrt{\Lambda}-\sqrt{\lambda})^2} + e^{-\frac{1}{2}(\sqrt{\Lambda}+\sqrt{\lambda})^2} \right]. \tag{3.40}$$

An equal-probability contour plot of the sampling distribution of the detection statistic is shown in Figure 9. The fact that we are able to write down *analytic* expressions for the sampling distributions for the detection statistic Λ is due to the simplicity of the signal and noise models. For more complicated real-world problems, these distributions would need to be generated *numerically* using fake signal injections and time-shifts to produce many different realizations of the data (signal plus noise) from which one can build up the distributions.

It is also important to point out that Λ is *not* a sufficient statistic for a, due to the fact that Λ involves the *square* of the maximum-likelihood estimate $\hat{a}$—i.e., $\Lambda = \hat{a}^2 N/\sigma^2$. Thus, we cannot take $p(\Lambda|a, \mathcal{M}_1)$ conditioned on Λ (assuming a flat prior on a from $[0, a_{\max}]$) to get the posterior distribution for a given d, since we would be missing out on data samples that give negative values for $\hat{a}$. Another way to see this is to start with $p(\Lambda|a, \mathcal{M}_1)$ given by (3.40), and then make a change of variables from Λ to $\hat{a}$ using the general transformation relation

$$p_Y(y)\,dy = p_X(x)\,dx \quad \Rightarrow \quad p_X(x) = \left[p_Y(y)\,|f'(x)|\right]_{y=f(x)}\,. \tag{3.41}$$

This leads to

$$\tilde{p}(\hat{a}|a, \mathcal{M}_1) = \frac{1}{\sqrt{2\pi}\sigma_{\hat{a}}} \left[e^{-\frac{1}{2\sigma_{\hat{a}}^2}(\hat{a}-a)^2} + e^{-\frac{1}{2\sigma_{\hat{a}}^2}(\hat{a}+a)^2} \right], \tag{3.42}$$

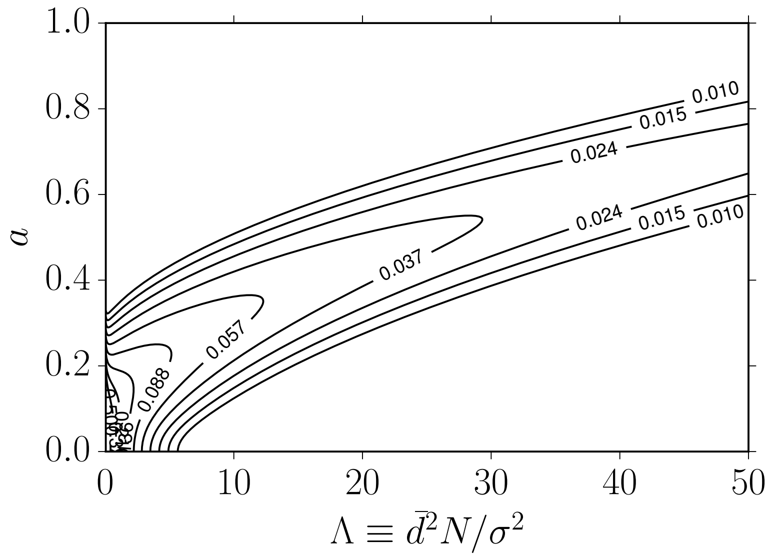


Figure 9: Equal-probability contour plot for the frequentist detection statistic $\Lambda \equiv \bar{d}^2 N/\sigma^2$ for a signal with amplitude $a \geq 0$.

which is properly normalized for $\hat{a} > 0$, but differs from (3.28) due to the second term involving $\hat{a} + a$. Thus, we need to construct $p(a|d)$ from (3.28)—and *not* from (3.42)—if we want the posterior to have the proper dependence on a.

3.5.1 Simulated data

For our example, we will take $N = 100$ samples, $\sigma = 1$, and $a_{\rm max} = 0.5$. We also simulate data with injected signals having amplitudes $a_0 = 0.05$ and 0.3, respectively. Since the expected signal-to-noise ratio, $a\sqrt{N}/\sigma$, is given by 0.5 and 3.0, these injections correspond to *weak* and (moderately) *strong* signals. Single realizations of the data for the two different injections are shown in Figure 10. The noise realization is the same for the two injections.

3.5.2 Frequentist analysis

Given the values for N, σ, and the probability distributions (3.39) and (3.40) for the frequentist detection statistic Λ, we can calculate the detection threshold for fixed false alarm probability α (which we will take to equal 10%), and the corresponding detection probability as a function of the amplitude a. The detection threshold turns out to equal $\Lambda_* = 2.9$ (so 10% of the area under the probability distribution $p(\Lambda|\mathcal{M}_0)$ is for $\Lambda \geq \Lambda_*$). The value of the amplitude a needed for 90% confidence detection probability with 10% false alarm probability is given by $a^{90\%,{\rm DP}} = 0.30$. (These results for the detection threshold and detection probability do *not* depend on the particular realizations of the simulated data.) The corresponding curves are shown in Figure 11.

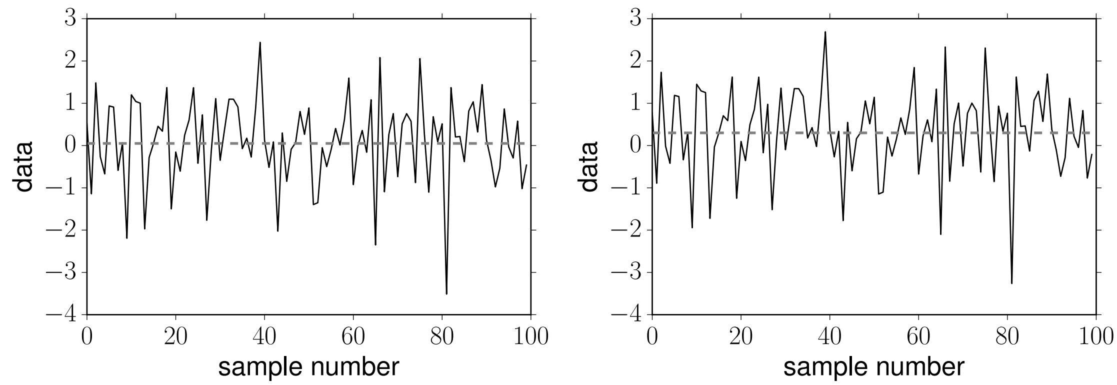


Figure 10: Examples of simulated data for weak (left panel) and strong (right panel) signals injected into the data—$a_0 = 0.05$ and 0.3, respectively.

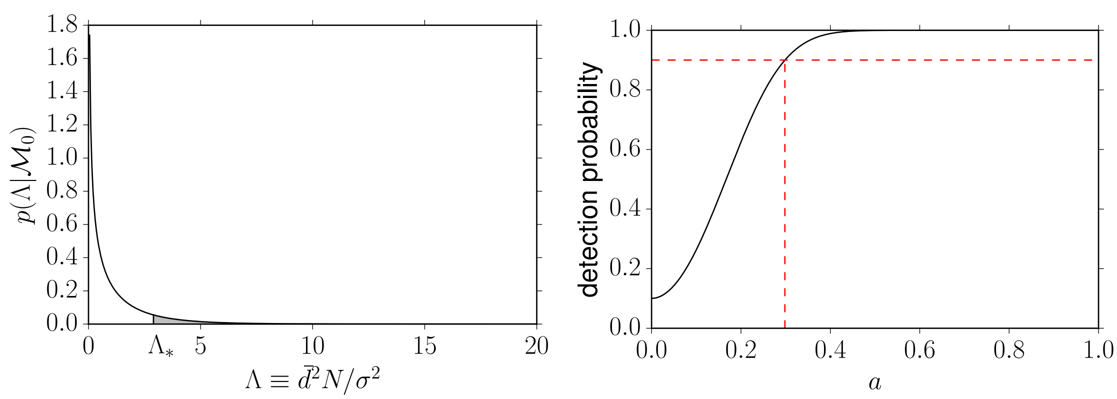


Figure 11: Left panel: Probability distribution for the frequentist detection statistic Λ for the noise-only model. The threshold value of the statistic for 10% false alarm probability is $\Lambda_* = 2.9$. Right panel: Detection probability as a function of the amplitude a. The value of the amplitude needed for 90% confidence detection probability with 10% false alarm probability is $a^{90\%,\mathrm{DP}} = 0.30$.

The sample mean of the data for the two simulations are given by $\bar{d} = 0.085$ and 0.335, respectively. Since $\hat{a} = \bar{d}$, these are also the values of the maximum-likelihood estimator of the amplitude a. The corresponding values of the detection statistic are $\Lambda_{\rm obs} = 0.72$ and 11.2 for the two injections, and have p-values equal to 0.45 and 9.0×10^{-4}, as shown in Figure 12. The 95% frequentist confidence interval is given simply by $[\hat{a} - 2\sigma_{\hat{a}}, \hat{a} + 2\sigma_{\hat{a}}]$, since $\hat{a}$ is Gaussian-distributed, and has values $[-0.11, 0.29]$ and $[0.14, 0.54]$, respectively. These intervals contain the true value of the amplitude for the two injections, $a_0 = 0.05$ and 0.3.

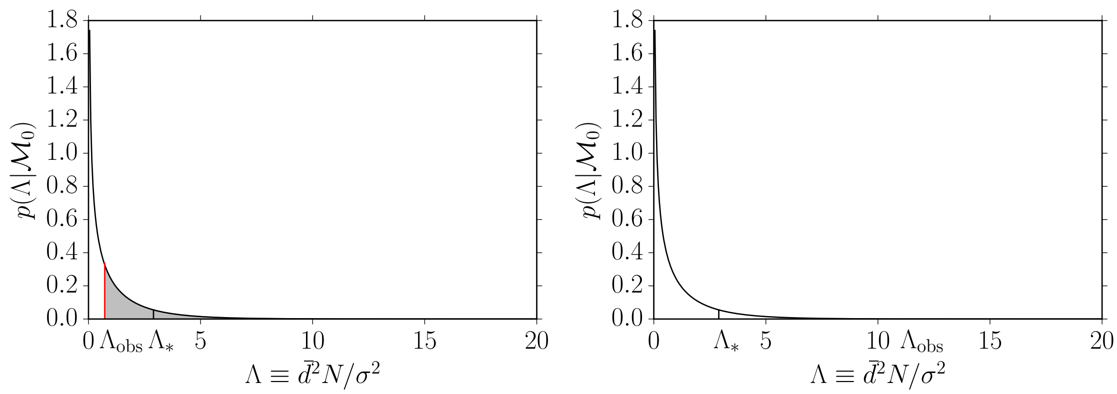


Figure 12: Graphical representation of the p-value calculation for the weak (left panel) and strong (right panel) injections. For the weak injection, $\Lambda_{\rm obs} = 0.72$ is marked by the red vertical line, with corresponding p-value 0.45. For the strong injection, $\Lambda_{\rm obs} = 11.2$ is sufficiently large that the corresponding red vertical line is not visible on this graph. The p-value for the strong injection is 9.0×10^{-4}.

The 90% confidence-level frequentist upper limits are $a^{90\%,{\rm UL}} = 0.20$ and 0.46, respectively. Figure 13 shows the probability distributions for the detection statistic Λ conditioned on these upper limit values for which the probability of obtaining $\Lambda \geq \Lambda_{\rm obs}$ is equal to 0.90.

3.5.3 Bayesian analysis

The results of the Bayesian analysis for the two different injections are summarized in Figure 14. The plots show the posterior distribution for the amplitude a given the value of the maximum-likelihood estimator $\hat{a}$, which (as we discussed earlier) is a sufficient statistic for the data d. Recall that the posterior for a for this example is simply a truncated Gaussian from 0 to $a_{\rm max}$ centered on $\hat{a}$, which could be negative, see (3.31). The left two panels show the graphical construction of the Bayesian 90% upper limit and 95% credible interval for the amplitude a for the weak injection, $a^{90\%,{\rm UL}} = 0.23$ and $[0, 0.26]$. The right two panels show similar plots for the strong injection, $a^{90\%,{\rm UL}} = 0.46$ and $[0.14, 0.54]$.

Finally, the Bayes factor for the signal-plus-noise model $\mathcal{M}_1$ relative to the noise-only

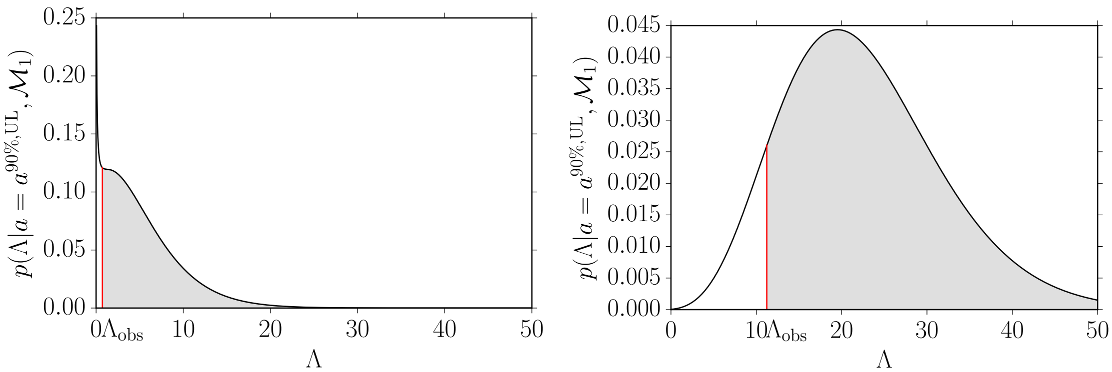


Figure 13: Probability distributions for the frequentist detection statistic Λ, conditioned on the value of the amplitude a for which the probability of obtaining $\Lambda \geq \Lambda_{obs}$ is equal to 0.90. These define the 90% confidence level frequentist upper limits $a^{90\%,UL} = 0.20$ and 0.46, respectively. The red vertical lines mark the value of Λ_{obs} for the weak (left panel, $\Lambda_{obs} = 0.72$) and strong (right panel, $\Lambda_{obs} = 11.2$) injections.

model $\mathcal{M}_0$ can be calculated by taking the ratio of the marginalized likelihood $p(d|\mathcal{M}_1)$ to $p(d|\mathcal{M}_0)$, where $p(d|\mathcal{M}_1)$ is obtained from (3.26) by marginalizing over $a \in [0, a_{max}]$. (There are no parameters to marginalize over for the noise-only model $\mathcal{M}_0$.) Doing this, we find $2 \ln \mathcal{B}_{10} = -2.2$ and 9.2 for the weak and strong signal injections, respectively. The Laplace approximation to this quantity, cf. (3.23), becomes

$$2 \ln \mathcal{B}_{10}(d) \simeq \frac{\hat{a}^2}{\sigma^2/N} + 2 \ln \left(\frac{\sqrt{2\pi}\sigma/\sqrt{N}}{a_{max}} \right) , \tag{3.43}$$

and gives -2.0 and 8.5, respectively.

3.5.4 Comparision summary

Table 4 summarizes the numerical results for the frequentist and Bayesian analyses. We see that the frequentist and Bayesian 90% upper limits and 95% intervals numerically agree for the strong injection, but differ slightly for the weak injection. The interpretation of these results is different, of course, for a frequentist and a Bayesian, given their different definitions of probability. But for a moderately strong signal in noisy data, we expect both approaches to yield a confident detection as they have for this simple example.

3.6 Likelihoods and priors for gravitational-wave searches

To conclude this section on statistical inference, we discuss some issues related to calculating the likelihood and choosing a prior in the context of searches for gravitational-wave signals using a network of gravitational-wave detectors.

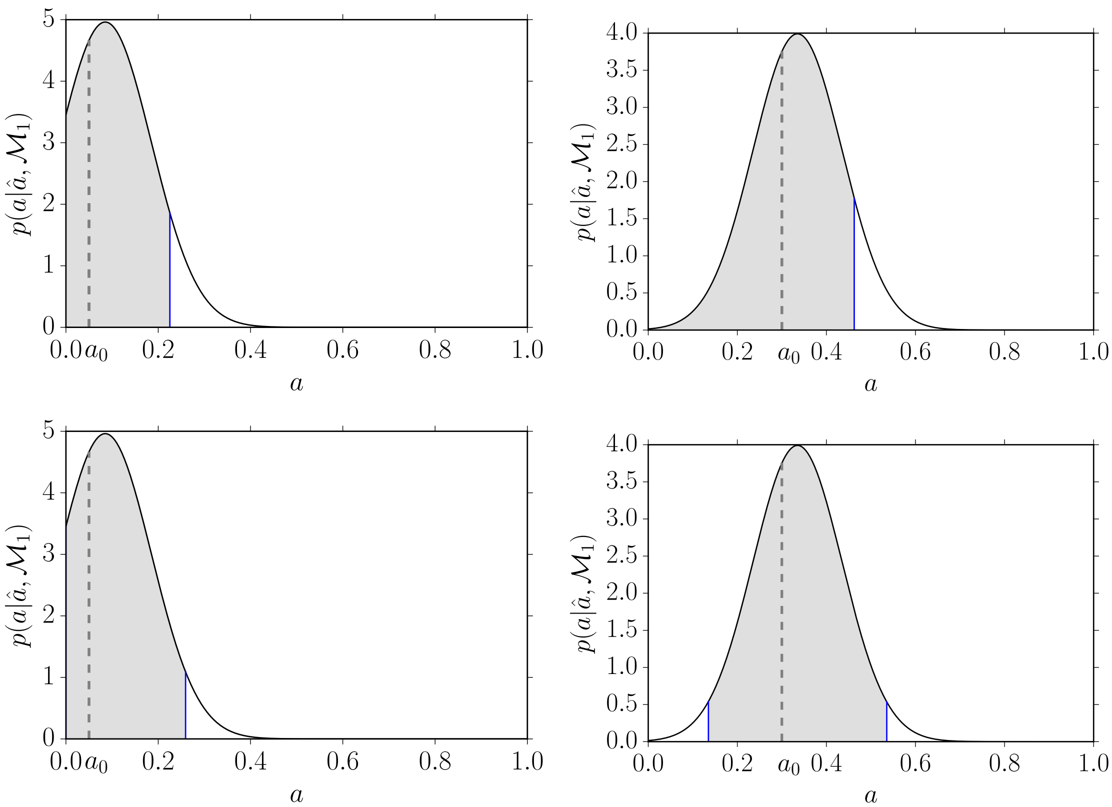


Figure 14: Posterior distributions for the amplitude a given the value of the maximum-likelihood estimator $\hat{a}$. The left two panels are for the weak injection; the right two panels are for the strong injection. The top two plots illustrate the graphical construction of Bayesian 90% upper limits for the two injections; the bottom two plots illustate the graphical construction of the Bayesian 95% credible intervals. The dashed vertical lines indicate the values of the injected signal amplitude a_0, which equal 0.05 and 0.3, respectively.

	(weak injection, $a_0 = -0.05$) Frequentist	Bayesian	(strong injection, $a_0 = 0.3$) Frequentist	Bayesian
Detection threshold (Λ_*)	2.9	—	2.9	—
Detection statistic ($\Lambda_{\rm obs}$)	0.72	—	11.2	—
p-value	0.45	—	9.0×10^{-4}	—
90% upper limit	0.20	0.23	0.46	0.46
95% interval	$[-0.11, 0.29]$	$[0, 0.26]$	$[0.14, 0.54]$	$[0.14, 0.54]$
ML estimator ($\hat{a}$)	0.085	0.085	0.335	0.335
Bayes factor ($2 \ln \mathcal{B}_{10}$)	—	−2.2	—	9.2
Laplace approximation	—	−2.0	—	8.5

Table 4: Tabular summary of the frequentist and Bayesian analysis results for the simulated data (both weak and strong injections). A dash indicates that a particular quantity is not relevant for either the frequentist or Bayesian analysis.

3.6.1 Calculating the likelihood

Defining the likelihood function (for either a frequentist or Bayesian analysis) involves understanding the instrument response and the instrument noise. The data collected by gravitational-wave detectors comes in a variety of forms. For ground-based interferometers such as LIGO and Virgo, the data comes from the error signal in the differential arm-length control system, which is non-linearly related to the laser phase difference, which in turn is linearly related to the gravitational-wave strain. For pulsar timing arrays, the data comes from the arrival times of radio pulses (derived from the folded pulse profiles), which must be corrected using a complicated timing model that takes into account the relative motion of the telescopes and the pulsars, along with the spin-down of the pulsars, in addition to a variety of propagation effects. The timing residuals formed by subtracting the timing model from the raw arrival times contain perturbations due to gravitational waves integrated along the line of sight to the pulsar. For future space-based gravitational-wave detectors such as eLISA, the data will be directly read out from phase meters that perform a heterodyne measurement of the laser phase. Synthetic combinations of these phase read outs (chosen to cancel laser phase noise) are then linearly proportional to the gravitational-wave strain.

Since gravitational waves can be treated as small perturbations to the background geometry, the time delays or laser phase/frequency shifts caused by a gravitational wave can easily be computed. These idealized calculations have then to be related to the actual observations, either by propagating the effects through an instrument response model, or, alternatively, inverting the response model to convert the measured data to something proportional to the gravitational-wave strain. (For example, most LIGO analyses work with the calibrated strain, rather than the raw differential error signal.) If we assume that the gravitational-wave signal and the instrument noise are linearly independent, then the data taken at time t can be written as

$$d(t) = h(t) + n(t)\,, \tag{3.44}$$

where $h(t)$ is shorthand for the gravitational-wave metric perturbations $h_{ab}(t, \vec{x})$ convolved with the instrument response function and converted into the appropriate quantity—phase shift, time delay, differential arm length error, etc. (A detailed calculation of $h(t)$ and the associated detector response functions will be given in Section 5.2.) As mentioned above, the data $d(t)$ may be the quantity that is measured directly, or, more commonly, some quantity that is derived from the measurements such as timing residuals or calibrated strain. In any analysis, it is important to marginalize over the model parameters used to make the conversion from the raw data.

The likelihood of observing $d(t)$ is found by demanding that the residual

$$r(t) \equiv d(t) - \bar{h}(t)\,, \tag{3.45}$$

be consistent with a draw from the noise distribution $p_n(x)$:

$$p(d(t)|\bar{h}(t)) = p_n(r(t)) = p_n(d(t) - \bar{h}(t))\,. \tag{3.46}$$

Here $\bar{h}(t)$ is our model[4] for the gravitational-wave signal. The likelihood of observing a collection of discretely-sampled data $d \equiv \{d_1, d_2, \cdots, d_N\}$, where $d_i \equiv d(t_i)$, is then given by $p(d|\bar{h}) = p_n(r)$, where $r \equiv \{r_1, r_2, \cdots, r_N\}$ with $r_i \equiv r(t_i)$. Since instrument noise is due to a large number of small disturbances combined with counting noise in the large-number limit, the central-limit theorem suggests that the noise distribution can be approximated by a multi-variate normal (Gaussian) distribution:

$$p(d|\bar{h}) = \frac{1}{\sqrt{\det(2\pi C_n)}}\, e^{-\frac{1}{2}\sum_{i,j} r_i \left(C_n^{-1}\right)_{ij} r_j}\,, \tag{3.47}$$

where C_n is the noise correlation matrix, with components

$$(C_n)_{ij} = \langle n_i n_j \rangle - \langle n_i \rangle \langle n_j \rangle\,. \tag{3.48}$$

If the noise is stationary, then the correlation matrix only depends on the lag $|t_i - t_j|$, and the matrix C_n can be diagonalized by transforming to the Fourier domain, where r_i should then be interpreted as $\tilde{r}(f_i)$. In practice the noise observed in most gravitational-wave experiments is neither stationary nor Gaussian (Section 9 and Appendix C), but (3.47) still serves as a good starting point for more sophisticated treatments. The Gaussian likelihood (3.47) immediately generalizes for a network of detectors:

$$p(d|\bar{h}) = \frac{1}{\sqrt{\det(2\pi C_n)}} e^{-\frac{1}{2}\sum_{Ii,Jj} r_{Ii} \left(C_n^{-1}\right)_{Ii,Jj} r_{Jj}}\,, \tag{3.49}$$

where I, J labels the detector, and i, j labels the discrete time or frequency sample for the corresponding detector. Note here that the parameters $\vec{\theta}$ appearing in (3.13) are the individual time or frequency samples $\bar{h}_i$.

[4] Since the model $\bar{h}(t)$ will most-likely differ from the actual $h(t)$, we use an overbar for the model to distinguish the two.

3.6.2 Choosing a prior

For Bayesian inference, it is also necessary to define a model $\mathcal{M}$ for the gravitational-wave signal, which is done by placing a prior $p(\bar{h}|\mathcal{M})$ on the samples $\bar{h}_i$. In some cases, a great deal is known about the signal model, such as when approximate solutions to Einstein's equations provide waveform templates. In that case the prior can be written as

$$p(\bar{h}|\mathcal{M}) = \delta(\bar{h} - \bar{h}(\vec{\theta}, \mathcal{M}))\, p(\vec{\theta}|\mathcal{M}) \,. \tag{3.50}$$

Marginalizing over $\bar{h}$ converts the posterior $p(\bar{h}|d)$ to a posterior distribution for the signal parameters $p(\vec{\theta}|d, \mathcal{M})$. In other cases, such as for short-duration bursts associated with certain violent astrophysical events, much less is known about the possible signals and weaker priors have to be used. Models using wavelets, which have finite time-frequency support, and priors that favor connected concentrations of power in the time-frequency plane are commonly used these "unmodeled burst" searches. At the other end of the spectrum from deterministic point sources are the statistically-isotropic stochastic backgrounds that are thought to be generated by various processes in the early universe, or through the superposition of a vast number of weak astrophysical sources. In the case of Gaussian stochastic signals, the prior for a signal $\bar{h} = (\bar{h}_+(\hat{n}), \bar{h}_\times(\hat{n}))$ coming from direction $\hat{n}$ direction has the form

$$p(\bar{h}|\mathcal{M}) = \frac{1}{2\pi S_h} e^{-(\bar{h}_+^2(\hat{n}) + \bar{h}_\times^2(\hat{n}))/2S_h} \,, \tag{3.51}$$

where S_h is the power spectrum of the background. As we shall show in Section 4, marginalizing over $\bar{h}$ converts the posterior $p(\bar{h}|d)$ to a posterior $p(S_h|d, \mathcal{M})$ for S_h.

4 Correlations

Correlation is not cause, it is just a 'music of chance'. *Siri Hustvedt*

Stochastic gravitational waves are indistinguishable from unidentified instrumetal noise in a single detector, but are correlated between pairs of detectors in ways that differ, in general, from instrumental noise. Cross-correlation methods basically use the random output of one detector as a template for the other, taking into account the physical separation and relative orientation of the two detectors. In this section, we introduce cross-correlation methods in the context of both frequentist and Bayesian inference, analyzing in detail a simple toy problem (the data are "white" and we ignore complications that come from the separation and relative orientation of the detectors—this we discuss in detail in Section 5). We also briefly discuss possible alternatives to cross-correlation methods, e.g., using a null channel as a noise calibrator.

The basic idea of using cross-correlation to search for stochastic gravitational-waves can be found in several early papers [78, 82, 119, 42, 43, 66]. The derivation of the likelihood function in Section 4.2 follows that of [49]; parts of Section 4.4 are also discussed in [24, 58].

4.1 Basic idea

The key property that allows one to distinguish a a stochastic gravitational-wave background from instrumental noise is that the gravitational-wave signal is correlated across multiple detectors while instrumental noise typically is not. To see this, consider the simplest possible example, i.e., a single sample of data from two colocated and coaligned detectors:

$$\begin{aligned} d_1 &= h + n_1\,, \\ d_2 &= h + n_2\,. \end{aligned} \tag{4.1}$$

Here h denotes the common gravitational-wave signal and n_1, n_2 the noise in the two detectors. To cross correlate the data, we simply form the product of the two samples, $\hat{C}_{12} \equiv d_1 d_2$. The expected value of the correlation is then

$$\langle \hat{C}_{12} \rangle = \langle d_1 d_2 \rangle = \langle h^2 \rangle + \langle n_1 n_2 \rangle + \cancelto{0}{\langle h n_2 \rangle} + \cancelto{0}{\langle n_1 h \rangle} = \langle h^2 \rangle + \langle n_1 n_2 \rangle\,, \tag{4.2}$$

since the gravitational-wave signal and the instrumental noise are uncorrelated. If the instrument noise in the two detectors are also uncorrelated, then

$$\langle n_1 n_2 \rangle = 0\,, \tag{4.3}$$

which implies

$$\langle \hat{C}_{12} \rangle = \langle h^2 \rangle \equiv S_h\,. \tag{4.4}$$

This is just the variance (or power) of the stochastic gravitational-wave signal. So by cross-correlating data in two (or more) detectors, we can extract the common gravitational-wave component.

We have assumed here that there is no cross-correlated noise (instrumental or environmental). If there is cross-correlated noise, then the simple procedure describe above needs to be augmented. This will be discussed in more detail in Section 9.5.

4.2 Relating correlations and likelihoods

The cross-correlation approach arises naturally from a standard likelihood analysis if we adopt a Gaussian stochastic template for the signal. Revisiting the example from the previous section, let's assume that the detector noise is Gaussian-distributed with variances S_{n_1} and S_{n_2}. Then the likelihood function for the data $d \equiv (d_1, d_2)$ for the noise-only model $\mathcal{M}_0$ is simply

$$p(d|S_{n_1}, S_{n_2}, \mathcal{M}_0) = \frac{1}{2\pi\sqrt{S_{n_1}S_{n_2}}}\, e^{-\frac{d_1^2}{2S_{n_1}}}\, e^{-\frac{d_2^2}{2S_{n_2}}} . \tag{4.5}$$

For the signal-plus-noise model $\mathcal{M}_1$, we have

$$p(d|S_{n_1}, S_{n_2}, \bar{h}, \mathcal{M}_1) = \frac{1}{2\pi\sqrt{S_{n_1}S_{n_2}}}\, e^{-\frac{(d_1-\bar{h})^2}{2S_{n_1}}}\, e^{-\frac{(d_2-\bar{h})^2}{2S_{n_2}}} , \tag{4.6}$$

where the gravitational wave signal $\bar{h}$ is assumed to be a Gaussian random deviate with probability distribution

$$p(\bar{h}|S_h, \mathcal{M}_1) = \frac{1}{\sqrt{2\pi S_h}}\, e^{-\frac{\bar{h}^2}{2S_h}} . \tag{4.7}$$

In most applications we are not interested in the value of $\bar{h}$, but rather the power S_h. Marginalizing over $\bar{h}$, the likelihood takes the form

$$p(d|S_{n_1}, S_{n_2}, S_h, \mathcal{M}_1) = \frac{1}{\sqrt{\det(2\pi C)}} e^{-\frac{1}{2}\sum_{I,J=1}^{2} d_I C_{IJ}^{-1} d_J} , \tag{4.8}$$

where

$$C = \begin{bmatrix} S_{n_1} + S_h & S_h \\ S_h & S_{n_2} + S_h \end{bmatrix} . \tag{4.9}$$

Maximizing the likelihood with respect to S_h, S_{n_1} and S_{n_2} yields the maximum-likelihood estimators

$$\begin{aligned} \hat{S}_h &= d_1 d_2 = \hat{C}_{12} , \\ \hat{S}_{n_1} &= d_1^2 - d_1 d_2 , \\ \hat{S}_{n_2} &= d_2^2 - d_1 d_2 . \end{aligned} \tag{4.10}$$

Thus, the cross-correlation statistic $\hat{C}_{12}$ is the maximum-likelihood estimator for a Gaussian stochastic gravitational wave template with zero mean and variance S_h.

4.3 Extension to multiple data samples

The extension to multiple data samples

$$\begin{aligned} d_{1i} &= h_i + n_{1i} , \quad i = 1, 2, \cdots, N , \\ d_{2i} &= h_i + n_{2i} , \quad i = 1, 2, \cdots, N , \end{aligned} \tag{4.11}$$

is fairly straightforward. In the following two subsections, we consider the cases where the detector noise and stochastic signal are either: (i) both *white* (i.e., the data are uncorrelated between time samples) or (ii) both *colored* (i.e., allowing for correlations in time). The white noise example will be analyzed in more detail in Sections 4.4–4.6.

4.3.1 White noise and signal

If the detector noise and stochastic signal are both white, then the likelihood functions for the data $d \equiv \{d_{1i}; d_{2i}\}$, are simply *products* of the likelihoods (4.6) and (4.8) for the individual data samples. We can write these product likelihoods as single multivariate Gaussian distributions:

$$p(d|S_{n_1}, S_{n_2}, \mathcal{M}_0) = \frac{1}{\sqrt{\det(2\pi C_n)}}\, e^{-\frac{1}{2}d^T C_n^{-1} d}\,, \tag{4.12}$$

$$p(d|S_{n_1}, S_{n_2}, S_h, \mathcal{M}_1) = \frac{1}{\sqrt{\det(2\pi C)}}\, e^{-\frac{1}{2}d^T C^{-1} d}\,, \tag{4.13}$$

where

$$C_n = \begin{bmatrix} S_{n_1}\,\mathbb{1}_{N\times N} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & S_{n_2}\,\mathbb{1}_{N\times N} \end{bmatrix}, \tag{4.14}$$

$$C = \begin{bmatrix} (S_{n_1}+S_h)\,\mathbb{1}_{N\times N} & S_h\,\mathbb{1}_{N\times N} \\ S_h\,\mathbb{1}_{N\times N} & (S_{n_2}+S_h)\,\mathbb{1}_{N\times N} \end{bmatrix}. \tag{4.15}$$

The arguments in the exponential have the form

$$d^T C_n^{-1} d = \sum_{I,J=1}^{2}\sum_{i,j=1}^{N} d_{Ii} C^{-1}_{n\,Ii,Jj} d_{Jj}\,, \tag{4.16}$$

and similarly for $d^T C^{-1} d$. The maximum-likelihood estimators for this case are:

$$\begin{aligned} \hat{S}_h &\equiv \frac{1}{N}\sum_{i=1}^{N} d_{1i}d_{2i}\,, \\ \hat{S}_{n_1} &\equiv \frac{1}{N}\sum_{i=1}^{N} d_{1i}^2 - \frac{1}{N}\sum_{i=1}^{N} d_{1i}d_{2i}\,, \\ \hat{S}_{n_2} &\equiv \frac{1}{N}\sum_{i=1}^{N} d_{2i}^2 - \frac{1}{N}\sum_{i=1}^{N} d_{1i}d_{2i}\,. \end{aligned} \tag{4.17}$$

Note that these are just averages of the single-datum estimators (4.10) over the N independent data samples.

A couple of remarks are in order: (i) It is easy to show that the expectation values of the estimators are the true values of the parameters S_h, S_{n_1}, S_{n_2}. It is also fairly straightforward to calculate the variances of the estimators. In particular,

$$\mathrm{Var}(\hat{S}_h) \equiv \langle \hat{S}_h^2\rangle - \langle \hat{S}_h\rangle^2 = \frac{1}{N}\left[S_{n_1}S_{n_2} + S_h(S_{n_1}+S_{n_2}) + 2S_h^2\right]. \tag{4.18}$$

Note that this expression reduces to $\mathrm{Var}(\hat{S}_h) \approx S_{n_1}S_{n_2}/N$ in the weak-signal limit, $S_h \ll S_{n_I}$, for $I = 1, 2$. (ii) If we simply maximized the likelihood with respect to variations of

S_h, treating the noise variances S_{n_1} and S_{n_2} as *known* parameters, then the frequentist estimator of S_h would also include *auto-correlation* terms for each detector:

$$\hat{S}_h = \frac{1}{(S_{n_1}+S_{n_2})^2}\left[2S_{n_1}S_{n_2}\frac{1}{N}\sum_{i=1}^{N} d_{1i}d_{2i} \right. \\ \left. +S_{n_2}\left(\frac{1}{N}\sum_{i=1}^{N} d_{1i}^2 - S_{n_1}\right) + S_{n_1}\left(\frac{1}{N}\sum_{i=1}^{N} d_{2i}^2 - S_{n_2}\right)\right]. \tag{4.19}$$

In practice, however, the noise variances are not known well enough to be able to extract useful information from the auto-correlation terms; they actually worsen the performance of the simple cross-correlation estimator when the uncertainty in S_{n_1} or S_{n_2} is greater than or equal to S_h.

4.3.2 Colored noise and signal

For the case where the detector noise and stochastic signal are colored, it simplest to work in the frequency domain, since the Fourier components are *independent* of one another. (This assumes that the data are *stationary*, so that there is no preferred origin of time.) Assuming multivariate Gaussian distributions as before, the variances S_{n_1}, S_{n_2}, and S_h generalize to *power spectral densitites*, which are functions of frequency defined by

$$\langle \tilde{n}_I(f)\tilde{n}_I^*(f')\rangle = \frac{1}{2}\delta(f-f')\,S_{n_I}(f)\,, \qquad \langle \tilde{h}(f)\tilde{h}^*(f')\rangle = \frac{1}{2}\delta(f-f')\,S_h(f)\,, \tag{4.20}$$

where $I=1,2$ and tilde denotes Fourier transform.[5] The factor of 1/2 in (4.20) is for *one-sided* power spectra, for which the integral of the power spectrum over *positive* frequencies equals the variance of the data:

$$\mathrm{Var}[h] = \int_0^\infty df\; S_h(f)\,. \tag{4.21}$$

For N samples of discretely-sampled data from each of two detectors $I=1,2$ (total duration T), the likelihood function for a Gaussian stochastic signal template becomes [23, 49]:

$$p(d|S_{n_1},S_{n_2},S_h,\mathcal{M}_1) = \prod_{k=0}^{N/2-1}\frac{1}{\det(2\pi\tilde{C}(f_k))}e^{-\frac{1}{2}\sum_{I,J}\tilde{d}_I^*(f_k)\tilde{C}_{IJ}^{-1}(f_k)\tilde{d}_J(f_k)}\,, \tag{4.22}$$

where

$$\tilde{C}(f) = \frac{T}{4}\begin{bmatrix} S_{n_1}(f)+S_h(f) & S_h(f) \\ S_h(f) & S_{n_2}(f)+S_h(f)\end{bmatrix}. \tag{4.23}$$

Here $k=0,1,\cdots,N/2-1$ labels the discrete positive frequencies. There is no square root of the determinant in the denominator of (4.22) since the volume element for the

[5]Our convention for Fourier transform is $\tilde{h}(f)=\int_{-\infty}^{\infty} dt\, e^{-i2\pi f t}h(t)$.

probability density involves both the real and imaginary parts of the Fourier transformed data.

We do not bother to write down the maximum-likelihood estimators of the signal and noise power spectral densities for this particular example. We will return to this problem in Section 6, where we discuss the *optimally-filtered* cross-correlation statistic for isotropic stochastic backgrounds. There one assumes a particular spectral *shape* for the gravitational-wave power spectral density, and then simply estimates its overall amplitude. That simplifies the analysis considerably.

4.4 Maximum-likelihood detection statistic

Let's return to the example discussed in Section 4.3.1, which consists of N samples of data in each of two detectors, having uncorrelated white noise and a common white stochastic signal. As described in Section 3.4, one can calculate a frequentist detection statistic based on the *maximum-likelihood ratio*:

$$\Lambda_{\rm ML}(d) \equiv \frac{\max_{S_{n_1},S_{n_2},S_h} p(d|S_{n_1},S_{n_2},S_h,\mathcal{M}_1)}{\max_{S_{n_1},S_{n_2}} p(d|S_{n_1},S_{n_2},\mathcal{M}_0)}\,. \tag{4.24}$$

Substituting (4.12) and (4.13) for the likelihood functions and performing the maximizations yields

$$\Lambda_{\rm ML}(d) = \left[1 - \frac{\hat{S}_h^2}{\hat{S}_1\hat{S}_2}\right]^{-N/2}\,, \tag{4.25}$$

where

$$\hat{S}_1 \equiv \frac{1}{N}\sum_{i=1}^N d_{1i}^2 = \hat{S}_{n_1} + \hat{S}_h\,, \qquad \hat{S}_2 \equiv \frac{1}{N}\sum_{i=1}^N d_{2i}^2 = \hat{S}_{n_2} + \hat{S}_h\,. \tag{4.26}$$

Note that the these estimators involve only *autocorrelations* of the data. In the absence of a signal, they are maximum-likelihood estimators of the noise variances S_{n_1} and S_{n_2}. But in the presence of a signal, they are maximum-likelihood estimators of the *combined* variances $S_1 \equiv S_{n_1} + S_h$ and $S_2 \equiv S_{n_2} + S_h$.

Recall that for comparison with Bayesian model selection calculations, it is convenient to define the frequentist statistic $\Lambda(d)$ as twice the logarithm of the maximum-likelihood ratio:

$$\Lambda(d) \equiv 2\ln\left(\Lambda_{\rm ML}(d)\right) = -N\ln\left[1 - \frac{\hat{S}_h^2}{\hat{S}_1\hat{S}_2}\right]\,. \tag{4.27}$$

In the limit that the stochastic gravitational-wave signal is *weak* compared to the detector noise—i.e., $S_h \ll S_{n_I}$, for $I = 1, 2$—the above expression reduces to

$$\Lambda(d) \simeq \frac{\hat{S}_h^2}{\hat{S}_1\hat{S}_2/N} \simeq \frac{\hat{S}_h^2}{\hat{S}_{n_1}\hat{S}_{n_2}/N}\,. \tag{4.28}$$

This is just the squared signal-to-noise ratio of the cross-correlation statistic. Note also that $\hat{S}_h^2/\hat{S}_1\hat{S}_2$ is the normalized cross-correlation (i.e., *coherence*) of the data from the two detectors. It is a measure of how well the data in detector 2 *matches* that in detector 1.

From (4.17), we see that $\Lambda(d)$ is a ratio of the square of a sum of products of Gaussian random variables to the product of a sum of squares of Gaussian random variables. This is a sufficiently complicated expression that we will estimate the distribution of $\Lambda(d)$ *numerically*, doing fake signal injections into many realizations of simulated noise to build up the sampling distribution. We do this explicitly in Section 4.6, when we compare the frequentist and Bayesian correlation methods for this example.

4.5 Bayesian correlation analysis

Compared to the frequentist cross-correlation analysis described above, a Bayesian analysis is conceptually much simpler. One simply needs the likelihood functions $p(d|S_{n_1}, S_{n_2}, \mathcal{M}_0)$ and $p(d|S_{n_1}, S_{n_2}, S_h, \mathcal{M}_1)$ given by (4.12) and (4.13), and joint prior probability distributions for the signal and noise parameters. For our example, we will assume that the signal and noise parameters are *statistically independent* of one another so that the joint prior distributions factorize into a product of priors for the individual parameters. We use Jeffrey's priors for the individual noise variances:

$$p_I(S_{n_I}) \propto 1/S_{n_I}\,, \qquad I = 1, 2\,, \tag{4.29}$$

and a flat[6] prior for the signal variance:

$$p(S_h) = \text{const}\,. \tag{4.30}$$

Then, using Bayes' theorem (3.13), we obtain the joint posterior distribution:

$$\begin{aligned} p(S_{n_1}, S_{n_2}, S_h|d, \mathcal{M}_1) &= \frac{p(d|S_{n_1}, S_{n_2}, S_h, \mathcal{M}_1)p(S_{n_1}, S_{n_2}, S_h|\mathcal{M}_1)}{p(d|\mathcal{M}_1)} \\ &\propto p(d|S_{n_1}, S_{n_2}, S_h, \mathcal{M}_1)\frac{1}{S_{n_1}}\frac{1}{S_{n_2}}\,, \end{aligned} \tag{4.31}$$

where $p(d|\mathcal{M}_1)$ is the evidence (or marginalized likelihood) for the signal-plus-noise model $\mathcal{M}_1$. (Similar expressions can be written down for the noise-only model $\mathcal{M}_0$.) The marginalized posterior distributions for the signal and noise parameters are given by marginalizing over the other parameters. For example,

$$p(S_h|d, \mathcal{M}_1) \propto \int \frac{dS_{n_1}}{S_{n_1}} \int \frac{dS_{n_2}}{S_{n_2}}\, p(d|S_{n_1}, S_{n_2}, S_h, \mathcal{M}_1) \tag{4.32}$$

for the signal variance S_h.

Correlations enter the Bayesian analysis via the covariance matrix C that appears in the likelihood function $p(d|S_{n_1}, S_{n_2}, S_h, \mathcal{M}_1)$. The covariance matrix for the data includes the cross-detector signal correlations, as we saw in (4.15). So although one does not explicitly construct a cross-correlation statistic in the Bayesian framework, cross correlations do play an important role in the calculations.

[6] A flat prior for S_h yields more conservative (i.e., larger) upper limits for S_h than a Jeffrey's prior, since there is more prior weight at larger values of S_h for a flat prior than for a Jeffrey's prior.

4.6 Comparing frequentist and Bayesian cross-correlation methods

To explicitly compare the frequentist and Bayesian methods for handling cross-correlations, we simulate data for the white noise, white signal example that we have been discussing in the previous subsections. The particular realization of data that we generate has $N = 100$ samples with $S_{n_1} = 1$, $S_{n_2} = 1.5$, and $S_h = 0.3$. Plots of the simulated data in the two detectors are given in Figure 15.

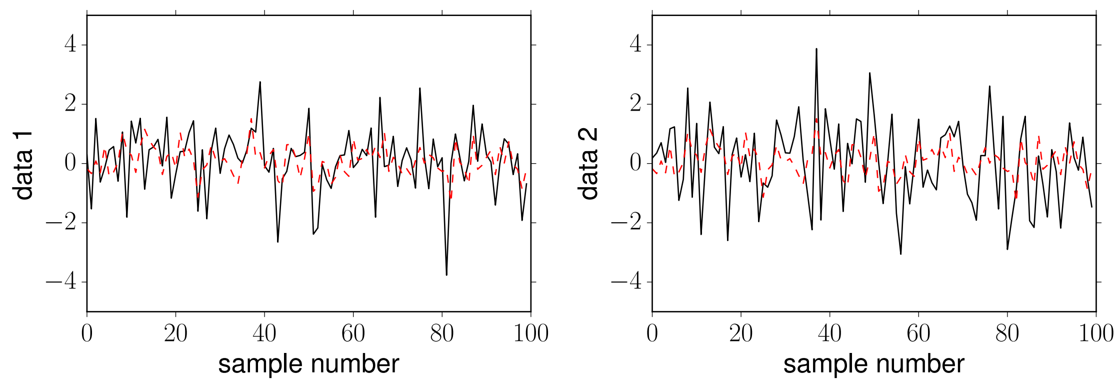


Figure 15: Simulated data in the two detectors. The detector output is shown by the black curves; the common stochastic signal is shown by the red dashed curves.

4.6.1 Frequentist analysis

The frequentist maximum-likelihood estimators (4.17) are very easy to calculate. For the simulated data they have values:

$$\hat{S}_{n_1} = 0.78\,, \quad \hat{S}_{n_2} = 1.46\,, \quad \hat{S}_h = 0.40\,. \tag{4.33}$$

In addition

$$\Lambda_{\rm ML}(d) = 44\,, \quad \Lambda(d) \equiv 2\ln\left(\Lambda_{\rm ML}(d)\right) = 7.6\,. \tag{4.34}$$

The weak-signal approximation to $\Lambda(d)$, given by (4.28), is significantly larger (having a value of 14), since the injected stochastic signal for this case was relatively strong, with the injected S_h equal to $0.3S_{n_1}$ and $0.2S_{n_2}$. In addition, for this realization of data, the signal variance was overestimated while both noise variances were underestimated, leading to a much larger value than the nominal squared signal-to-noise ratio of 6.

As mentioned previously, the form (4.27) of the detection statistic $\Lambda(d)$ is sufficiently complicated that it was simplest to resort to numerical simulations to estimate its sampling distribution, $p(\Lambda|S_{n_1}, S_{n_2}, S_h, \mathcal{M}_1)$. We took 50 values for each of S_{n_1}, S_{n_2}, and S_h in the interval $[0, 3]$, and then for each of the corresponding 50^3 points in parameter space, we generated 10^4 realizations of the data, yielding 10^4 values of $\Lambda(d)$. By histogramming these values for each point in parameter space, we were able to estimate the probability density function (and also the cumulative distribution function) for Λ.

Figure 16 shows the frequentist 90% confidence-level *exclusion* and *inclusion* regions for our simulated data with $\Lambda_{\rm obs} = 7.6$. The 90% confidence-level exclusion region $\mathcal{E}_{90\%}$ lies *above* the red surface; it consists of points (S_{n_1}, S_{n_2}, S_h) satisfying

$$\text{Prob}\,(\Lambda \geq \Lambda_{\rm obs} | (S_{n_1}, S_{n_2}, S_h) \in \mathcal{E}_{90\%}) \geq 0.90\,. \tag{4.35}$$

The region *below* the red surface is the 90% confidence-level inclusion region $\mathcal{I}_{90\%}$. Note that construction of these regions is such that the *true* values of the parameters S_{n_1}, S_{n_2}, and S_h have a 90% frequentist probability of lying in $\mathcal{I}_{90\%}$. This generalizes, to multiple parameters, the definition of the frequentist 90% confidence-level upper-limit for a single parameter, which was discussed in detail in Section 3.2.3. Note that it is not correct to simply "cut" the surface using the maximum-likelihood point estimates $\hat{S}_{n_1} = 0.78$ and $\hat{S}_{n_2} = 1.46$ to obtain a single value for $S_h^{90\%,{\rm UL}}$. One needs to include the whole region in order to get the correct frequentist coverage.

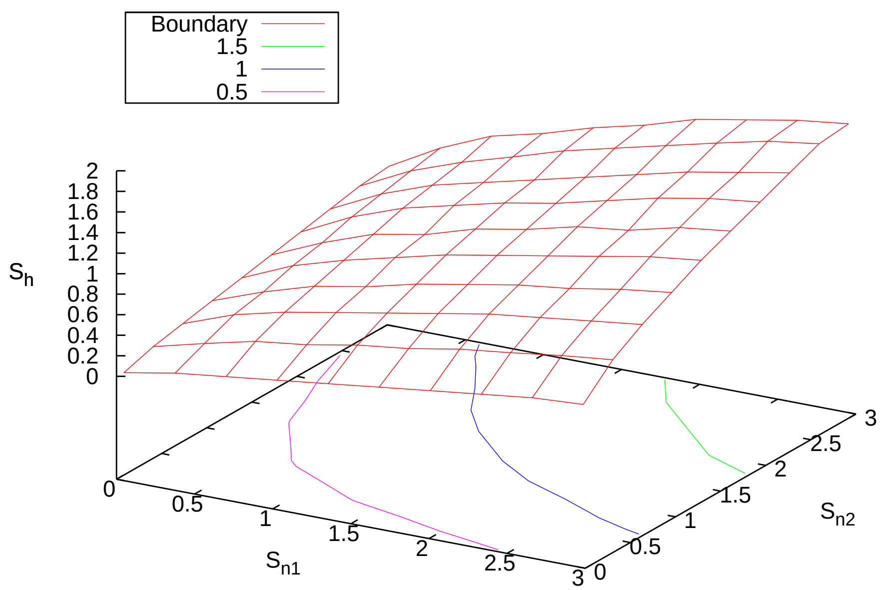


Figure 16: Frequentist 90% confidence level exclusion and inclusion regions for the simulated data with $\Lambda_{\rm obs} = 7.6$. The 90% exclusion region $\mathcal{E}_{90\%}$ lies above the red surface; the 90% inclusion region $\mathcal{I}_{90\%}$ lies below the red surface. The green, blue and magenta curves are projections of the $S_h = 1.5, 1.0, 0.5$ level surfaces of the boundary onto the (S_{n_1}, S_{n_2}) plane.

A similar procedure can be used to estimate sampling distributions for the frequentist maximum-likelihood estimators $\hat{S}_{n_1}$, $\hat{S}_{n_2}$, and $\hat{S}_h$. From these distributions, one can then calculate e.g., frequentist 95% confidence-level exclusion and inclusion *regions* for the given point estimates. For example, $(S_{n_1}, S_{n_2}, S_h) \in \mathcal{I}_{95\%}$ for the observed point estimate $\hat{S}_{h,\text{obs}}$ if and only if $\hat{S}_{h,\text{obs}}$ is contained in the symmetric 95% confidence interval centered on the mode of the probability distribution $p(\hat{S}_h|S_{n_1}, S_{n_2}, S_h, \mathcal{M}_1)$. These regions again generalize to multiple parameters the definition of a frequentist confidence *interval* for a single parameter, which was discussed in detail in Section 3.2.4. They will be different, in general, for the different maximum-likelihood estimators. But in order to move on to the Bayesian analysis for this example, we will leave the explicit construction of these regions to the interested reader.

4.6.2 Bayesian analysis

For the Bayesian analysis of this example, we limit ourselves to calculating the Bayes factor $2\ln\mathcal{B}_{10}(d)$ comparing the noise-only and signal-plus-noise models $\mathcal{M}_0$ and $\mathcal{M}_1$, as well as the posterior distributions for the three parameters S_h, S_{n_1}, and S_{n_2}. Following the procedure described above in Section 4.5 we find, for this particular realization of data,

$$\mathcal{B}_{10} = 10\,, \quad 2\ln\mathcal{B}_{10}(d) = 4.6\,. \tag{4.36}$$

This Bayes factor corresponds to *positive* evidence (see Table 3) in favor of a correlated stochastic signal in the data.

Figure 17 shows the marginalized posterior $p(S_h|d, \mathcal{M}_1)$ for the stochastic signal variance given the data d and signal-plus-noise model $\mathcal{M}_1$. The peak of the posterior lies close the frequentist maximum-likelihood estimator $\hat{S}_h = 0.40$ (blue dotted vertical line), and easily contains the injected value in its 95% Bayesian credible interval (grey shaded region). Figure 18 shows similar plots for the marginalized posteriors for the noise variances S_{n_1} and S_{n_2} for both the signal-plus noise model $\mathcal{M}_1$ (blue curves) and the noise-only model $\mathcal{M}_0$ (green curves). For comparison, the frequentist maximum-likelihood estimators $\hat{S}_{n_1}, \hat{S}_{n_2} = 0.78, 1.46$ and $1.18, 1.86$ for the two models are shown by the corresponding (blue and green) dotted vertical lines. Again, the peaks of the Bayesian posterior distributions lie close to these values. The 95% Bayesian credible intervals for S_{n_1} and S_{n_2} for the signal-plus-noise model $\mathcal{M}_1$ are also shown (grey shaded region). These interval easily contain the injected values for these two parameters.

4.7 What to do when cross-correlation methods aren't available

Cross-correlation methods can be used whenever one has two or more detectors that respond to a common gravitational-wave signal. The beauty of such methods is that even though a stochastic background is another source of "noise" in a single detector, one can effectively use the output of one detector as a *template* for the output of the other detector. This is similar to matched-filtering for a deterministic signal in a single detector [83, 195]. But with only a single detector, searches for a stochastic background need some other way to distinguish the signal from the noise—e.g., a difference between the spectra of the noise and the gravitational-wave signal, or the modulation of an anisotropic signal due to the motion of the detector (as is expected for the confusion-noise from galactic compact white dwarf binaries for LISA). Without some way of differentiating instrumental noise from gravitational-wave "noise", there is no hope of detecting a stochastic background.

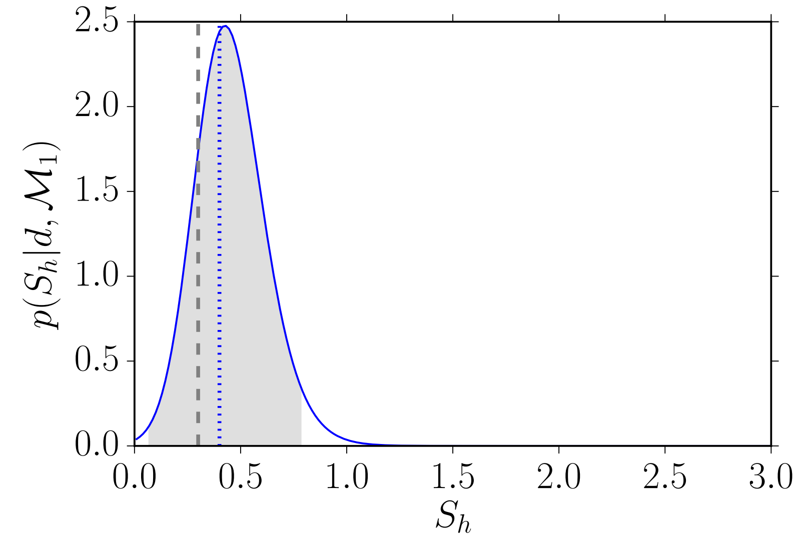


Figure 17: Marginalized posterior distribution for the stochastic signal variance S_h for the signal-plus-noise model $\mathcal{M}_1$. The actual value of S_h used for the simulation is shown by grey dashed vertical line. The 95% Bayesian credible interval centered on the mode of the distribution is the grey-shaded region. For comparison, the frequentist maximum-likelihood estimator of S_h is shown by the blue dotted vertical line.

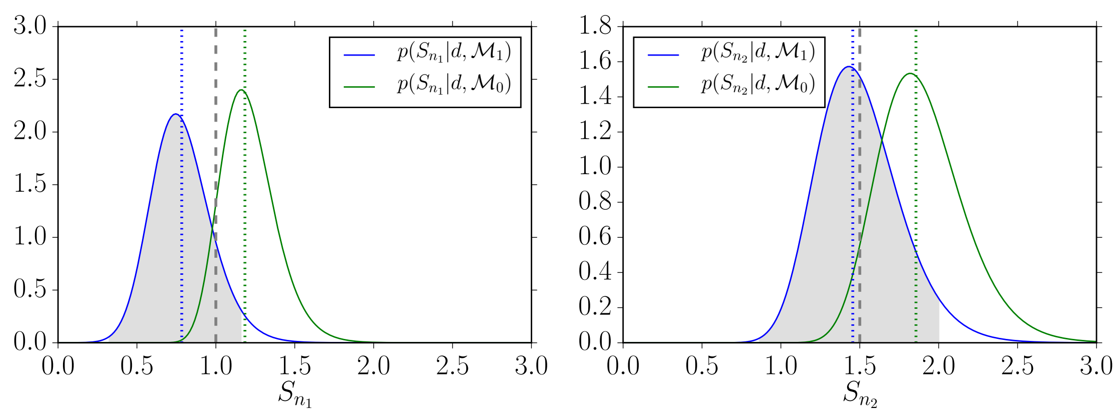


Figure 18: Marginalized posterior distributions for the detector noise variances S_{n_1} (left panel) and S_{n_2} (right panel) for the signal-plus-noise model $\mathcal{M}_1$ (blue curves) and the noise-only model $\mathcal{M}_0$ (green curves), respectively. The actual values of S_{n_1} and S_{n_2} used for the simulation are shown by grey dashed vertical lines. The 95% Bayesian credible intervals for the signal-plus-noise model $\mathcal{M}_1$ are the grey-shaded regions. For comparison, the frequentist estimators of S_{n_1} and S_{n_2} for the two models are shown by the (blue and green) dotted vertical lines.

As a simple example, suppose that we have N samples of data from each of two detectors $I = 1, 2$ (which we will call *channels* in what follows), but let's assume that the second channel is *insensitive* to the gravitational-wave signal:

$$\begin{aligned} d_{1i} &= h_i + n_{1i}\,, \qquad && i = 1, 2, \cdots, N\,, \\ d_{2i} &= n_{2i}\,, \qquad && i = 1, 2, \cdots, N\,. \end{aligned} \tag{4.37}$$

Then if we make the same assumptions as before for the signal and the noise, it follows that the likelihood function for the data $d \equiv \{d_{1i}; d_{2i}\}$ is given by

$$p(d|S_{n_1}, S_{n_2}, S_h, \mathcal{M}_1) = \frac{1}{\sqrt{\det(2\pi C)}} e^{-\frac{1}{2} d^T C^{-1} d}\,, \tag{4.38}$$

where

$$C = \begin{bmatrix} (S_{n_1} + S_h)\, \mathbb{1}_{N\times N} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & S_{n_2}\, \mathbb{1}_{N\times N} \end{bmatrix} \tag{4.39}$$

is the covariance matrix of the data. Since the off-diagonal blocks of the covariance matrix are identically zero, it is clear that we will not be able to use the cross-correlation methods developed in the previous sections. So we need to do something else if we are going to extract the gravitational-wave signal from the noise.

4.7.1 Single-detector excess power statistic

If we knew S_{n_1} a priori, then we could construct an *excess power* statistic from the auto-correlated data to estimate the signal variance:

$$\hat{S}_h \equiv \frac{1}{N} \sum_{i=1}^{N} d_{1i}^2 - S_{n_1}\,. \tag{4.40}$$

(This is effectively how Penzias and Wilson discovered the CMB [133]; they observed excess antenna noise that they couldn't attribute to any other known source of noise.) But as mentioned at the end of Section 4.3.1, typically we do not know the detector noise well enough to use such a statistic, since the uncertainty in S_{n_1} is much greater than the variance of the gravitational-wave signal that we are trying to detect. This is definitely the case for ground-based detectors like LIGO, Virgo, etc.. An exception to this "rule" will probably be the predicted *foreground* signal from galactic white-dwarf binaries in the LISA band. For frequencies below a few mHz, the gravitational-wave confusion noise from these binaries is expected to dominate the LISA instrument noise [85, 37, 84, 128].

4.7.2 Null channel method

If it were possible to make an *off-source* measurement using detector 1, then we could estimate the noise variance S_{n_1} directly from the detector output, free of contamination from gravitational waves. Using this noise estimate, $\hat{S}_{n_1}$, we could then define our excess power statistic as

$$\hat{S}_h \equiv \frac{1}{N} \sum_{i=1}^{N} d_{1i}^2 - \hat{S}_{n_1}\,. \tag{4.41}$$

Unfortunately, such off-source measurements are not possible, since you cannot shield a gravitational-wave detector from gravitational waves. However, in certain cases one can construct a particular *combination* of the data (called a *null channel*) for which the response to gravitational waves is strongly suppressed. The *symmetrized Sagnac* combination of the data for LISA [181, 87] is one such example.

So let's assume that channel 2 for our example is such a null channel, and let's also assume that there is some relationship between the noise in the two channels—e.g., $S_{n_1} = aS_{n_2}$, with $a > 0$. (For colored noise, the variances would be replaced by power spectra and a would be replaced by a function of frequency—i.e., a *transfer function* relating the noise in the two channels.) To begin with, we will also assume that a is *known*. Then the data from the second channel can be used as a *noise calibrator* for the first channel. The frequentist estimators for this scenario are:

$$\begin{aligned}
\hat{S}_{n_2} &= \frac{1}{N}\sum_{i=1}^{N} d_{2i}^2\,, \\
\hat{S}_{n_1} &= a\hat{S}_{n_2}\,, \\
\hat{S}_h &= \frac{1}{N}\sum_{i=1}^{N} d_{1i}^2 - \hat{S}_{n_1}\,.
\end{aligned} \tag{4.42}$$

These are the maximum-likelihood estimators of the signal and noise parameters, derived from the likelihood (4.38) with S_{n_1} replaced by aS_{n_2}. In the Bayesian framework, the relation $S_{n_1} = aS_{n_2}$ is encoded in the joint prior probability distribution

$$p(S_{n_1}, S_{n_2}) = \delta(S_{n_1} - aS_{n_2})p_2(S_{n_2})\,, \tag{4.43}$$

which eliminates S_{n_1} as an independent variable. The marginalized posterior distribution for the signal variance S_h, assuming a flat prior $p_h(S_h) = \text{const}$, is then

$$p(S_h|d) \propto \int dS_{n_2}\, p(d|S_{n_1} = aS_{n_2}, S_{n_2}, S_h)p_2(S_{n_2})\,. \tag{4.44}$$

In the more realistic scenario where the transfer function a is not known a priori, but is described by its own prior probability distribution $p_a(a)$, we have

$$p(S_{n_1}, S_{n_2}, a) = \delta(S_{n_1} - aS_{n_2})p_a(a)p_2(S_{n_2}) \tag{4.45}$$

and

$$p(S_h|d) \propto \int da \int dS_{n_2}\, p(d|S_{n_1} = aS_{n_2}, S_{n_2}, S_h)p_a(a)p_2(S_{n_2})\,. \tag{4.46}$$

This integral can be done numerically given priors for S_{n_2} and a.

To help illustrate the above discussion, Figure 19 shows plots of several different posterior distributions for S_h, corresponding to different choices for the prior distribution $p_a(a)$. For these plots, we chose a *Jeffrey's prior* for S_{n_2}:

$$p_2(S_{n_2}) \propto 1/S_{n_2}\,, \tag{4.47}$$

and a *log-normal* prior for a:

$$p(a|\mu,\sigma) = \frac{1}{a}\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(\ln a-\mu)^2}{\sigma^2}} \,. \tag{4.48}$$

The different curves correspond to different values of μ and σ:

$$\begin{aligned} \mu &\equiv \ln A\,, \qquad A = a_0, 0.67a_0, 1.5a_0\,,\\ \sigma &\equiv \ln\Sigma\,, \qquad \Sigma = 1, 1.1, 1.25, 1.5, 2\,, \end{aligned} \tag{4.49}$$

where a_0 denotes the nominal (true) value of a. Note that $A = 0.67a_0$ and $1.5a_0$ correspond to priors for a that are *biased* away from its true value $a = a_0$. In addition, 68% of the prior distribution is contained in the region $a \in [A/\Sigma, A\Sigma]$ (so $\Sigma = 1$ corresponds to a delta-function prior—i.e., no uncertainty in a). The particular realization that we used consisted of $N = 100$ samples of data (4.37) with $S_h = 1$, $S_{n_2} = 1$, and $S_{n_1} = a_0 S_{n_2}$ with $a_0 = 1$. Note that for the biased priors for a (associated with the dashed and dotted curves in Figure 19), an under (over) estimate in a corresponds to over (under) estimate in S_h, as S_h is effectively the difference between the estimated variance in channel 1 and a times the estimated variance in channel 2. For this particular realization of the data, the mode of the "0%, unbiased" posterior for S_h is about 20% less than the injected value, $S_h = 1$. On average, they would agree.

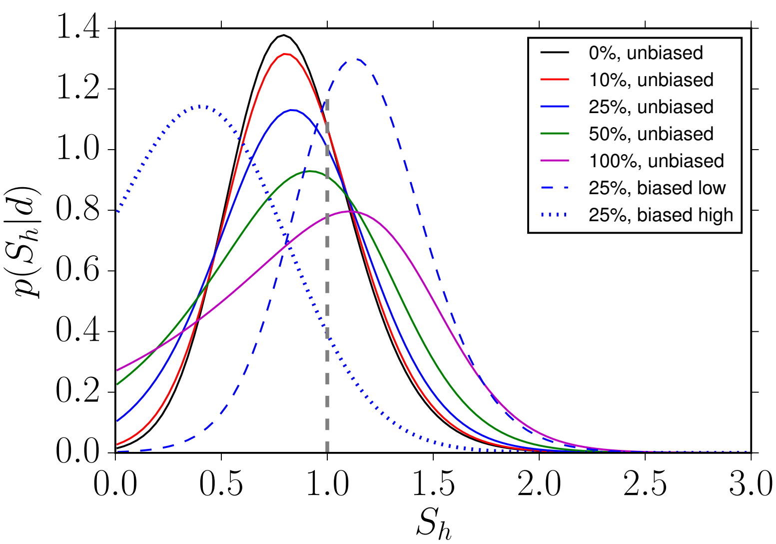


Figure 19: Posterior distributions for S_h for the null channel analysis, corresponding to different priors for the parameter a, which relates the instrumental noise variances in the two channels. The labels "p%, unbiased" correspond to $A = a_0$ and $\Sigma = 1 + p/100$; the labels "25%, biased low (or high)" correspond to $A = 0.67a_0$ (or $1.5a_0$) and $\Sigma = 1.25$. The vertical grey dashed line corresponds to the injected value of S_h.

5 Geometrical factors

> There is geometry in the humming of the strings, there is music in the spacing of the spheres. *Pythagoras*

In the previous sections, we ignored many details regarding detector response and detector geometry. We basically assumed that the detectors were *isotropic*, responding equally well to all gravitational waves, regardless of the waves' directions of propagation, frequency content, and polarization. We also ignored any loss in sensitivity in the correlations between data from two or more detectors, due to the separation and relative orientation of the detectors. But these details *are* important if we want to design optimal (or near-optimal) data analysis algorithms to search for gravitational waves. To specify the likelihood function, for example, requires models not only for the gravitational-wave signal and instrument noise, but also for the reponse of the detectors to the waves that a source produces.

In this section, we fill in these details. We first discuss the response of a single detector to an incident gravitational wave. We then show how these non-trivial detector responses manifest themselves in the correlation between data from two or more detectors. The results are first derived in a general setting making no assumption, for example, about the wavelength of a gravitational wave to the characteristic size of a detector. The general results are then specialized, as appropriate, to the case of ground-based and space-based laser interferometers, spacecraft Doppler tracking, and pular timing arrays. We conclude this section by discussing how the motion of a detector relative to the gravitational-wave source affects the detector response.

The approach we take in this section is similar in spirit to that of [81], attempting to unify the treatment of detector response functions and correlation functions across different gravitational-wave detectors. Readers interested in more details about the effect of detector geometry on the correlation of data from two or more detectors should see the original papers by Hellings and Downs [82] for pulsar timing arrays, and Flanagan [66] and Christensen [42, 43] for ground-based laser interferometers.

5.1 Detector response

Gravitational waves are time-varying perturbations to the background geometry of space-time. Since gravitational waves induce time-varying changes in the separation between two freely-falling objects (so-called test masses), gravitational-wave detectors are designed to be as sensitive as possible to this changing separation. For example, a resonant bar detector acts like a giant tuning fork, which is set into oscillation when a gravitational wave of the natural frequency of the bar is incident upon it. These oscillations produce a stress against the equilibrium electromagnetic forces that exist within the bar. The stress (or oscillation) is measured by a strain gauge (or accelerometer), indicating the presence of a gravitational wave. The response for a bar detector is thus the fractional change in length of the bar, $h(t) = \Delta l(t)/l$, induced by the wave. Since the length of the bar is typically much smaller than the wavelength of a gravitational wave at the bar's resonant frequency, the response is most easily computed using the geodesic deviation equation [121] for the time-varying tidal field.

In this article, we will focus our attention on *beam* detectors, which use electromagnetic radiation to monitor the separation of two or more freely-falling objects. Spacecraft Doppler tracking, pulsar timing arrays, and ground- and space-based laser interferometers (e.g., LIGO-like and LISA-like detectors) are all examples of beam detectors, which can be used to search for gravitational waves (see, e.g., Section 4.2 in [147]).

5.1.1 Spacecraft Doppler tracking

For spacecraft Doppler tracking, pulses of electromagnetic radiation are sent from one test mass (e.g., a radio transmitting tower on Earth) to another (e.g., the Cassini probe), and then bounced back (or coherently transponded) from the second test mass to the first. From the arrival times of the returning pulses, one can calculate the fractional change in the frequency of the emitted pulses induced by a gravitational wave. The detector response for such a measurement is thus

$$h_{\rm doppler}(t) \equiv \frac{\Delta\nu(t)}{\nu_0} = \frac{d\Delta T(t)}{dt}\,, \tag{5.1}$$

where $\Delta T(t)$ is the deviation of the round-trip travel time of a pulse away from the value it would have had at time t in the absence of the gravitational wave. A schematic representation of $\Delta T(t)$ for spacecraft Doppler tracking is given in Figure 20.

5.1.2 Pulsar timing

Pulsar timing is even simpler in the sense that we only have *one-way* transmission of electromagnetic radiation (i.e., radio pulses are emitted by a pulsar and received by a radio receiver on Earth). The response for such a system is simply the timing residual

$$h_{\rm timing}(t) = \Delta T(t)\,, \tag{5.2}$$

which is the difference between the measured time of arrival of a radio pulse and the expected time of arrival of the pulse (as determined from a detailed timing model for the pulsar) due to the presence of a gravitational wave. A schematic representation of $\Delta T(t)$ for a pulsar timing measurement is given in Figure 21.

5.1.3 Laser interferometers

For laser interferometers like LIGO or LISA, the detector response is the phase difference in the laser light sent down and back the two arms of the interferometer. Again, the phase difference can be calculated in terms of the change in the round-trip travel time of the laser light from one test mass (e.g., the beam splitter) to another (e.g., one of the end test masses). If we consider an equal-arm Michelson interferometer with unit vectors $\hat{u}$ and $\hat{v}$ pointing from the beam splitter to the end masses in each of the arms, then

$$h_{\rm phase}(t) \equiv \Delta\Phi(t) = 2\pi\nu_0\Delta T(t)\,, \tag{5.3}$$

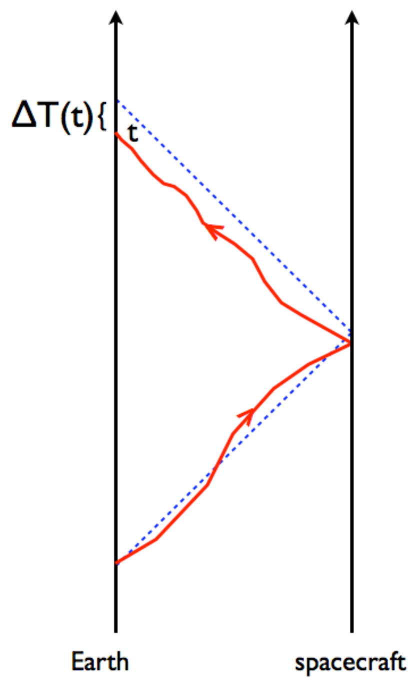


Figure 20: A space-time diagram representation of $\Delta T(t)$ for a two-way spacecraft Doppler tracking measurement. Time increases vertically upward. The vertical arrows are space-time worldlines for the Earth and a spacecraft. The measurement is made at time t. The blue dotted line shows the trajectory of a pulse of electromagnetic radiation in the absence of a gravitational wave; the red solid line shows the trajectory in the presence of a gravitational wave.

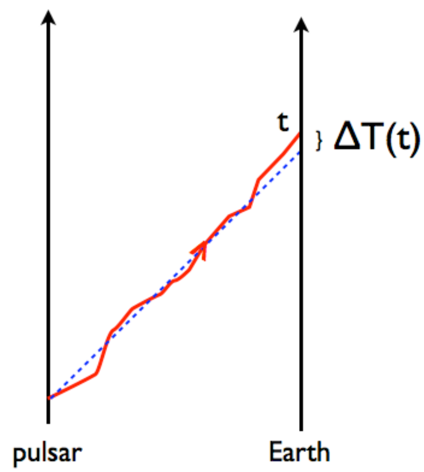


Figure 21: A space-time diagram representation of $\Delta T(t)$ for a (one-way) pulsar timing residual measurement. Time increases vertically upward. The vertical arrows are space-time worldlines for a pulsar and a detector on Earth. The measurement is made at time t. The blue dotted line shows the trajectory of the radio pulse in the absence of a gravitational wave; the red solid line shows the trajectory in the presence of a gravitational wave.

where $\Delta T(t) \equiv T_{\hat{u},\mathrm{rt}}(t) - T_{\hat{v},\mathrm{rt}}(t)$ is the difference of the round-trip travel times, and ν_0 is the frequency of the laser light. (See Figure 22.) Alternatively, one often writes the interferometer response as a *strain* measurement in the two arms

$$h_{\mathrm{strain}}(t) \equiv \frac{\Delta L(t)}{L} = \frac{\Delta T(t)}{2L/c}\,, \tag{5.4}$$

where $\Delta L(t) \equiv L_{\hat{u}}(t) - L_{\hat{v}}(t)$ is the difference of the proper lengths of the two arms (having unperturbed length L), and $\Delta T(t)$ is the difference in round-trip travel times as before. Thus, interferometer phase and strain response are simply related to one another.

Calculation of $\Delta T(t)$ for beam detectors is most simply carried out in the transverse-traceless gauge[7] [121, 148, 80] since the unperturbed separation L of the two test masses can be large or comparable to the wavelength $\lambda \equiv c/f$ of an incident gravitational wave having frequency f. This is definitely the case for pulsar timing where L is of order a few kpc, and for spacecraft Doppler tracking where L is of order tens of AU. It is also the case for space-based detectors like LISA ($L = 5 \times 10^6$ km) for gravitational waves with frequencies around a tenth of a Hz. On the other hand, for Earth-based detectors like LIGO ($L = 4$ km), $L \ll \lambda$ is a good approximation below a few kHz. Thus, the approach that we will take in the following subsections is to calculate the detector response in general, not making any approximation *a priori* regarding the relative sizes of $\lambda = c/f$

[7] See [53, 103] for an alternative derivation of the response of a detector to gravitational waves, which is done in terms of the curvature tensor and not the metric perturbations.

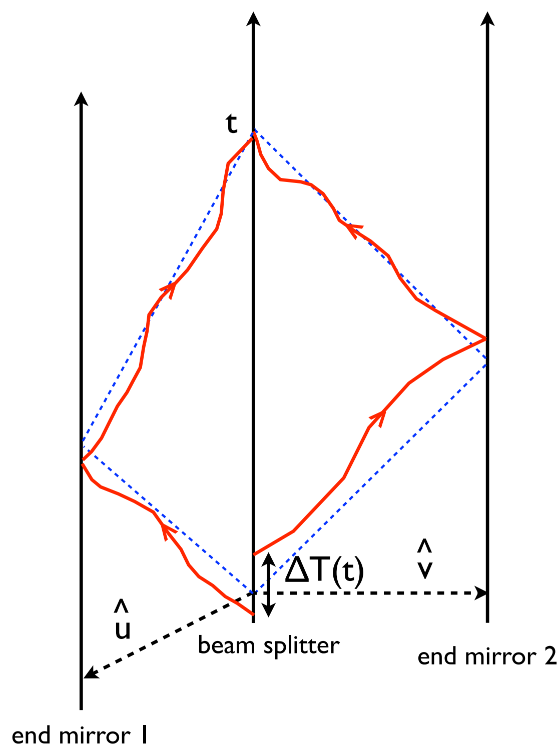


Figure 22: A space-time diagram representation of $\Delta T(t)$ for an equal-arm Michelson interferometer. Time increases vertically upward. The vertical arrows are space-time worldlines for the beam splitter and two end mirrors. The blue dotted lines show the trajectory of the laser light in the two arms of the interferometer in the absence of a gravitational wave; the red solid lines show the trajectory in the presence of a gravitational wave. The black dotted arrows, labeled $\hat{u}$ and $\hat{v}$, show the orientation of the two arms, from beam splitter to end mirrors, at $t = 0$, assuming an opening angle of $90°$.

and L. To recover the standard expressions (i.e., in the long-wavelength or small-antenna limit) for Earth-based detectors like LIGO will be a simple matter of taking the limit fL/c to zero. For reference, Table 5 summarizes the characteristic properties (i.e., size, characteristic frequency, sensitivity band, etc.) of different beam detectors.

Beam detector	L (km)	f_* (Hz)	f (Hz)	f/f_*	Relation
Ground-based interferometer	~ 1	$\sim 10^5$	$10 - 10^4$	$10^{-4} - 10^{-1}$	$f \ll f_*$
Space-based interferometer	$\sim 10^6$	$\sim 10^{-1}$	$10^{-4} - 10^{-1}$	$10^{-3} - 1$	$f \lesssim f_*$
Spacecraft Doppler tracking	$\sim 10^9$	$\sim 10^{-4}$	$10^{-6} - 10^{-3}$	$10^{-2} - 10$	$f \sim f_*$
Pulsar timing	$\sim 10^{17}$	$\sim 10^{-12}$	$10^{-9} - 10^{-7}$	$10^3 - 10^5$	$f \gg f_*$

Table 5: Characteristic properties of different beam detectors: column 2 is the arm length or characteristic size of the detector (tens of AU for spacecraft Doppler tracking; a few kpc for pulsar timing); column 3 is the frequency corresponding to the characteristic size of the detector, $f_* \equiv c/L$; columns 4 and 5 are the frequencies at which the detector is sensitive in units of Hz and units of f_*, respectively; and column 6 is the relationship between f and f_*.

5.2 Calculation of response functions and antenna patterns

Gravitational waves are weak. Thus, the detector response is *linear* in the metric perturbations $h_{ab}(t,\vec{x})$ describing the wave, and can be written as the convolution of the metric perturbations $h_{ab}(t,\vec{x})$ with the *impulse response* $R^{ab}(t,\vec{x})$ of the detector

$$h(t) = (\mathbf{R} * \mathbf{h})(t,\vec{x}) \equiv \int_{-\infty}^{\infty} d\tau \int d^3y\, R^{ab}(\tau,\vec{y}) h_{ab}(t-\tau,\vec{x}-\vec{y})\,, \tag{5.5}$$

where $\vec{x}$ is the location of the measurement at time t. In terms of a plane-wave expansion (2.1) of the metric perturbations, we have

$$h(t) = \int_{-\infty}^{\infty} df \int d^2\Omega_{\hat{n}} R^{ab}(f,\hat{n}) h_{ab}(f,\hat{n}) e^{i2\pi f t}\,, \tag{5.6}$$

or, in the frequency domain,

$$\tilde{h}(f) = \int d^2\Omega_{\hat{n}} R^{ab}(f,\hat{n}) h_{ab}(f,\hat{n})\,, \tag{5.7}$$

where[8]

$$R^{ab}(f,\hat{n}) = e^{i2\pi f \hat{n}\cdot\vec{x}/c} \int_{-\infty}^{\infty} \mathrm{d}\tau \int \mathrm{d}^3y\; R^{ab}(\tau,\vec{y})\, e^{-i2\pi f(\tau + \hat{n}\cdot\vec{y}/c)}\,. \tag{5.8}$$

[8]Some authors [42, 43, 66, 26, 47, 65], including us in the past, have defined the response function $R^{ab}(f,\hat{n})$ *without* the factor of $e^{i2\pi f\hat{n}\cdot\vec{x}/c}$. If one chooses coordinates so that the measurement is made at $\vec{x} = \vec{0}$, then these two definitions agree. Just be aware of this possible difference when reading the literature. To distinguish the two definitions, we will use the symbol $\bar{R}^{ab}(f,\hat{n})$ to denote the expression without the exponential term, i.e., $R^{ab}(f,\hat{n}) = e^{i2\pi f\hat{n}\cdot\vec{x}/c}\bar{R}^{ab}(f,\hat{n})$.

Further specification of the response function depends on the choice of gravitational-wave detector as well as on the basis tensors used to expand $h_{ab}(f,\hat{n})$, as we shall see below and in the following subsections.

For example, if we work in the polarization basis, with expansion coefficients $h_A(f,\hat{n})$, where $A=\{+,\times\}$, then

$$\tilde{h}(f)=\int d^2\Omega_{\hat{n}}\sum_A R^A(f,\hat{n})h_A(f,\hat{n})\,,\tag{5.9}$$

with

$$R^A(f,\hat{n})=R^{ab}(f,\hat{n})e^A_{ab}(\hat{n})\,.\tag{5.10}$$

If we work instead in the tensor spherical harmonic basis, with expansion coefficients $a^P_{(lm)}(f)$, where $P=\{G,C\}$, then

$$\tilde{h}(f)=\sum_{(lm)}\sum_P R^P_{(lm)}(f)a^P_{(lm)}(f)\,,\tag{5.11}$$

with

$$R^P_{(lm)}(f)=\int d^2\Omega_{\hat{n}}\, R^{ab}(f,\hat{n})Y^P_{(lm)ab}(\hat{n})\,.\tag{5.12}$$

Note that in the polarization basis the response function $R^A(f,\hat{n})$ is the detector response to a sinusoidal plane-wave with frequency f, coming from direction $\hat{n}$, and having polarization $A=+,\times$. Plots of $|R^A(f,\hat{n})|$ for fixed frequency f are *antenna beam patterns* for gravitational waves with polarization A. A plot of

$$\mathcal{R}(f,\hat{n})\equiv\left(|R^+(f,\hat{n})|^2+|R^\times(f,\hat{n})|^2\right)^{1/2}\tag{5.13}$$

for fixed frequency f is the beam pattern for an *unpolarized* gravitational wave—i.e., a wave having statistically equivalent $+$ and $\times$ polarization components.

Since the previous subsection showed that the response of all beam detectors can be written rather simply in terms of the change in the light-travel time of an electromagnetic wave propagating between two test masses, we now calculate $\Delta T(t)$ in various scenarios and use the resulting expressions to read-off the response functions $R^{ab}(f,\hat{n})$ for the different detectors. We also make plots of various antenna patterns.

5.2.1 One-way tracking

Consider two test masses located at position vectors $\vec{r}_1$ and $\vec{r}_2=\vec{r}_1+L\hat{u}$, respectively, in the presence of a plane gravitational wave propagating in direction $\hat{k}=-\hat{n}$, as shown in Figure 23. Then the change in the light-travel time for a photon emitted at $\vec{r}_1$ and received at $\vec{r}_2$ at time t is given by [63]:

$$\Delta T(t)=\frac{1}{2c}u^au^b\int_{s=0}^{L}ds\, h_{ab}(t(s),\vec{x}(s))\,,\tag{5.14}$$

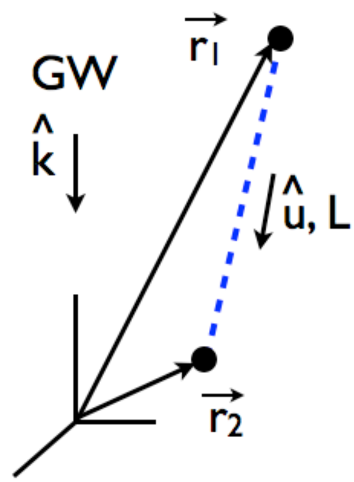


Figure 23: Geometry for calculating the change in the photon propagation time from $\vec{r}_1$ to $\vec{r}_2 = r_1 + L\hat{u}$ in the presence of a plane gravitational wave propagating in direction $\hat{k}$.

where the 0th-order expression for the photon trajectory can be used in h_{ab}:

$$t(s) = (t - L/c) + s/c\,, \quad \vec{x}(s) = \vec{r}_1 + s\hat{u}\,. \tag{5.15}$$

Since $h_{ab}(t, \vec{x}) = h_{ab}(t + \hat{n} \cdot \vec{x}/c)$ for a plane wave, it is relatively easy to do the integral. The result is

$$\begin{aligned}
\Delta T(t) &= \int_{-\infty}^{\infty} df \frac{1}{2} u^a u^b h_{ab}(f, \hat{n}) \\
&\qquad \frac{1}{i2\pi f} \frac{1}{1 + \hat{n} \cdot \hat{u}} \left[e^{i2\pi f(t_2 + \hat{n}\cdot\vec{r}_2/c)} - e^{i2\pi f(t_1 + \hat{n}\cdot\vec{r}_1/c)} \right] \qquad (5.16) \\
&= \int_{-\infty}^{\infty} df \frac{1}{2} u^a u^b h_{ab}(f, \hat{n})\, e^{i2\pi f(t + \hat{n}\cdot\vec{r}_2/c)} \\
&\qquad \frac{1}{i2\pi f} \frac{1}{1 + \hat{n} \cdot \hat{u}} \left[1 - e^{-\frac{i2\pi f L}{c}(1 + \hat{n}\cdot\hat{u})} \right]\,, \qquad (5.17)
\end{aligned}$$

where we factored out $e^{i2\pi f(t + \hat{n}\cdot\vec{r}_2/c)}$, corresponding to the time and location of the measurement, to get the last line. Note that the two terms in square brackets in (5.16) correspond to sampling the gravitational-wave phase at photon reception (location $\vec{r}_2$ at time $t_2 \equiv t$) and photon emission (location $\vec{r}_1$ at time $t_1 \equiv t - L/c$), respectively. In the context of pulsar timing, these two terms are called the *Earth term* and *pulsar term*, respectively.

From Equation (5.17), we can read-off the response function for a timing residual measurement, $h_{\rm timing}(t) \equiv \Delta T(t)$. It is

$$R^{ab}_{\rm timing}(f, \hat{n}) = \frac{1}{2} u^a u^b\, \mathcal{T}_{\vec{u}}(f, \hat{n} \cdot \hat{u}) e^{i2\pi f \hat{n}\cdot\vec{r}_2/c}\,, \tag{5.18}$$

where

$$\mathcal{T}_{\vec{u}}(f, \hat{n}\cdot\hat{u}) \equiv \frac{1}{i2\pi f}\,\frac{1}{1+\hat{n}\cdot\hat{u}}\,\left[1 - e^{-\frac{i2\pi f L}{c}(1+\hat{n}\cdot\hat{u})}\right] \tag{5.19}$$

$$= \frac{L}{c} e^{-\frac{i\pi f L}{c}(1+\hat{n}\cdot\hat{u})}\,\mathrm{sinc}\left(\frac{\pi f L}{c}[1+\hat{n}\cdot\hat{u}]\right) \tag{5.20}$$

is the *timing transfer function* for one-way photon propagation along $\vec{u} = L\hat{u}$. (Here $\mathrm{sinc}\,x \equiv \sin x/x$.) If we choose $\vec{r}_2$ to be the origin of coordinates, then $\mathcal{T}_{\vec{u}}(f, \hat{n}\cdot\hat{u})$ contains all the frequency-dependence of the timing response. For example, for normal incidence of the gravitational wave ($\hat{n}\cdot\hat{u} = 0$), $|\mathcal{T}_{\vec{u}}(f, 0)| = (L/c)\,|\mathrm{sinc}(\pi f L/c)|$. Figure 24 is a plot of $|\mathcal{T}_{\vec{u}}(f, 0)|$ versus frequency on a logarithmic frequency scale.

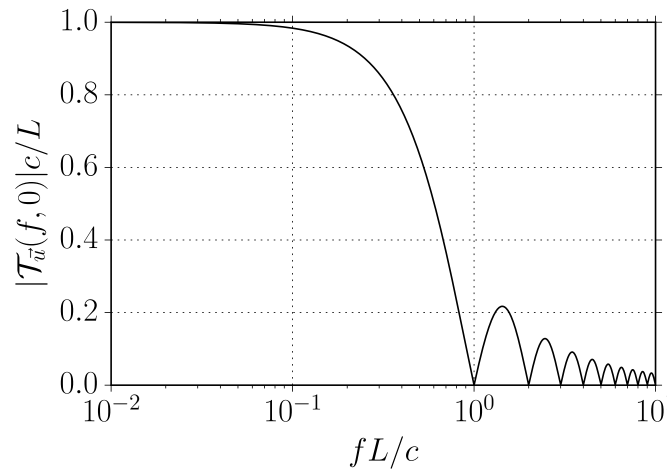


Figure 24: Magnitude of the one-way tracking timing transfer funtion $|\mathcal{T}_{\vec{u}}(f, 0)|$ for normal incidence of the gravitational wave, plotted on a logarithmic frequency scale. Nulls in the transfer function occur at frequencies equal to integer multiples of c/L.

If we choose instead to measure the fractional Doppler frequency shift of the incoming photons, then we need to differentiate the timing response with respect to t as indicated in (5.1). This simply pulls-down a factor of $i2\pi f$ from the exponential in $\Delta T(t)$, leading to

$$R^{ab}_{\mathrm{doppler}}(f, \hat{n}) = i2\pi f\, R^{ab}_{\mathrm{timing}}(f, \hat{n})\,. \tag{5.21}$$

Thus, the frequency-dependence of the Doppler frequency response is $i2\pi f$ times the timing transfer function $\mathcal{T}_{\vec{u}}(f, \hat{n}\cdot\hat{u})$. All of the above remarks are relevant for pulsar timing and *one-way* spacecraft Doppler tracking.

In Figure 25 we plot the antenna beam pattern (5.13) for unpolarized gravitational waves for a one-way tracking Doppler frequency measurement (e.g., pulsar timing) with $\hat{u} = -\hat{z}$. For this calculation, we chose $\vec{r}_2 = 0$ and ignored the exponential (i.e., 'pulsar') term in the timing transfer function, which yields

$$R^{A}_{\mathrm{doppler}}(f, \hat{n}) = \frac{1}{2}\frac{u^a u^b}{1+\hat{u}\cdot\hat{n}} e^{A}_{ab}(\hat{n}) \qquad \text{(Earth term only)}\,, \tag{5.22}$$

for the $A = +, \times$ polarization modes. Setting $\hat{u} = -\hat{z}$ and taking the gravitational waves to propagate inward (toward the origin), we find

$$\mathcal{R}_{\rm doppler}(\hat{n}) = \frac{1}{2}(1 + \cos\theta)\,, \tag{5.23}$$

which is axially symmetric around $\hat{u}$. The response is maximum when the photon and the gravitational wave both propagate in the same direction. Figure 26 shows plots of

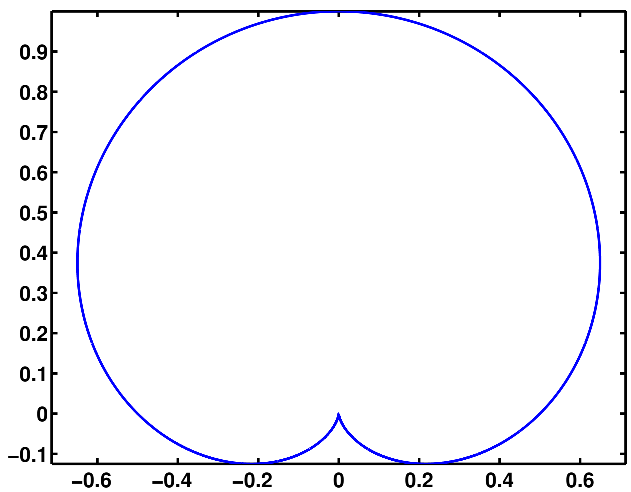


Figure 25: Antenna pattern for unpolarized gravitational waves for a one-way tracking Doppler frequency measurement with $\hat{u} = -\hat{z}$. The gravitational waves propagate toward the origin. The 3-d antenna pattern is axially symmetric around $\hat{u}$.

the real parts of the individual polarization basis response functions (5.22), represented as color bar plots on a Mollweide projection of the sky. For this plot we chose the pulsar to be located in the direction $(\theta, \phi) = (50°, 60°)$. (The direction $\hat{p}$ to the pulsar is given by $\hat{p} = -\hat{u}$.) The imaginary parts of both response functions are identically zero, so are not shown in the figure.

Making the same approximations as above, we can also calculate the corresponding Doppler-frequency response functions for the gradient and curl tensor spherical harmonic components $\{a^G_{lm}(f), a^C_{lm}(f)\}$ by performing the integration in (5.12). As shown in [69], this leads to[9]

$$R^G_{(lm)}(f) = 2\pi\, {}^{(2)}N_l Y_{lm}(\hat{p})\,, \qquad R^C_{(lm)}(f) = 0\,, \tag{5.24}$$

where ${}^{(2)}N_l$ is given by (2.8) and $\hat{p} = -\hat{u}$ is the direction on the sky to a pulsar. Note, somewhat surprisingly, that the curl response is *identically zero*. We will discuss the consequences of this result in more detail in Section 7.5.6, in the context of phase-coherent mapping of anisotropic gravitational-wave backgrounds.

[9] There is a factor of $(-1)^l$ difference between $R^G_{(lm)}(f)$ in (5.24) and (92) in [69]. The difference is due to the change in expressing response functions in terms of the direction to the gravitational-wave source, $\hat{n}$, as opposed to the direction of gravitational-wave propagation, $\hat{k} = -\hat{n}$. Appendix G provides expressions relating the response functions calculated using these two different conventions.

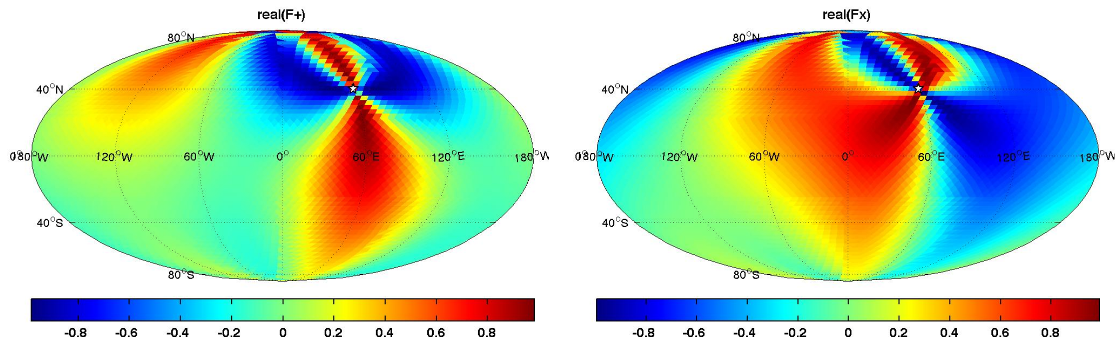


Figure 26: Mollweide projections of the repsonse functions $R^{+}_{\rm doppler}(\hat{n})$, $R^{\times}_{\rm doppler}(\hat{n})$, for one-way tracking Doppler frequency measurements corresponding to a pulsar located in the direction of the white star $(\theta, \phi) = (50^\circ, 60^\circ)$. The imaginary parts of both response functions are identically zero, so are not shown above.

5.2.2 Two-way tracking

To calculate $\Delta T(t)$ for *two-way* spacecraft Doppler tracking, we need to generalize the calculation of the previous subsection to include a return trip of the photon from $\vec{r}_2$ back to $\vec{r}_1$. This can be done by simply summing the expressions for the one-way timing residuals:

$$\Delta T(t) = \Delta T_{12}(t - L/c) + \Delta T_{21}(t) \tag{5.25}$$

where the subscripts on the ΔT's on the right-hand side of the above equation indicate the direction of one-way photon propagation (e.g., 12 indicates photon propagation from test mass 1 to test mass 2), and the arguments of ΔT_{12} and ΔT_{21} indicate when the photon arrived at test mass 2 and test mass 1, respectively. Doing this calculation leads to the following expression for the timing residual:

$$\begin{aligned}\Delta T(t) = \int_{-\infty}^{\infty} df \frac{1}{2} u^a u^b h_{ab}(f, \hat{n}) \, \frac{1}{i2\pi f} \Bigg[& \frac{1}{1 - \hat{n}\cdot\hat{u}} \, e^{i2\pi f(t + \hat{n}\cdot\vec{r}_1/c)} \\ & - \frac{2\hat{n}\cdot\hat{u}}{1 - (\hat{n}\cdot\hat{u})^2} \, e^{i2\pi f(t - L/c + \hat{n}\cdot\vec{r}_2/c)} - \frac{1}{1 + \hat{n}\cdot\hat{u}} \, e^{i2\pi f(t - 2L/c + \hat{n}\cdot\vec{r}_1/c)} \Bigg], \end{aligned} \tag{5.26}$$

which has *three* terms corresponding to the final reception of the photon at $\vec{r}_1$ at time t, the reflection of the photon at $\vec{r}_2$ at time $t - L/c$, and the emission of the photon at $\vec{r}_1$ at time $t - 2L/c$. The timing response function is given by

$$R^{ab}_{\rm timing}(f, \hat{n}) = \frac{1}{2} u^a u^b \, \mathcal{T}_{\vec{u},\rm rt}(f, \hat{n}\cdot\hat{u}) e^{i2\pi f \hat{n}\cdot\vec{r}_1/c} \,, \tag{5.27}$$

where

$$\begin{aligned}\mathcal{T}_{\vec{u},\rm rt}(f, \hat{n}\cdot\hat{u}) \equiv \frac{L}{c} e^{-\frac{i2\pi f L}{c}} \Bigg[& e^{-\frac{i\pi f L}{c}(1 - \hat{n}\cdot\hat{u})} \, {\rm sinc}\left(\frac{\pi f L}{c}[1 + \hat{n}\cdot\hat{u}]\right) \\ & + e^{\frac{i\pi f L}{c}(1 + \hat{n}\cdot\hat{u})} \, {\rm sinc}\left(\frac{\pi f L}{c}[1 - \hat{n}\cdot\hat{u}]\right) \Bigg] \end{aligned} \tag{5.28}$$

is the timing transfer function for *two-way* (or roundtrip) photon propagation along $\vec{u}$ and back. For normal incidence, the magnitude of the timing transfer function is given by $|\mathcal{T}_{\vec{u},\mathrm{rt}}(f,0)| = (2L/c)|\mathrm{sinc}(2\pi f L/c)|$, which is identical to the expression for one-way tracking with L/c replaced by $2L/c$. We also note that if we choose the origin of coordinates to be at $\vec{r}_1$ (which we can always do for a single detector), and if the frequency f is such that $fL/c \ll 1$, then the timing response simplifies to

$$R^{ab}_{\mathrm{timing}}(f,\hat{n}) = u^a u^b \, \frac{L}{c} \qquad (\text{for } fL/c \ll 1)\,. \tag{5.29}$$

We will use the terminology *small-antenna limit* (instead of *long-wavelength limit*) for this type of limit, since it avoids an ambiguity that might arise if we want to compare three or more length scales. For example, if we have two detectors that are physically separated and the wavelength of a gravitational wave is *large* compared to the size of each detector but *small* compared to the separation of the detectors, we would be in the long-wavelength limit with respect to detector size but in the short-wavelength limit with respect to detector separation. (This is actually the case for the current network of ground-based interferometers.) The terminology *small-antenna, large-separation limit* is more appropriate for this case.

5.2.3 Michelson interferometer

For an equal-arm Michelson interferometer, the timing residual that we calculate is the difference in the round-trip light-travel times down and back each of the arms. (See Figure 27.) If we let $\vec{u}$ and $\vec{v}$ denote the vectors pointing from e.g., the beam splitter

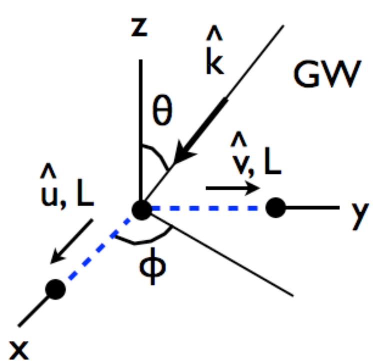


Figure 27: Geometry for calculating the difference in the round-trip light-travel times in the two arms of a Michelson interferometer: $\hat{k} = -\hat{n}$ is the direction of propagation for a plane gravitational wave; $\hat{u}$ and $\hat{v}$ are unit vectors that point from the vertex of the interferometer (e.g., the beam splitter) to the two end masses; and L denotes the lengths of each of the arms in the absence of a gravitational wave.

to the two end mirrors for LIGO, or from one spacecraft to the other two spacecraft for

LISA, then[10]

$$\Delta T(t) \equiv T_{\vec{u},\mathrm{rt}}(t) - T_{\vec{v},\mathrm{rt}}(t) = \Delta T_{\vec{u},\mathrm{rt}}(t) - \Delta T_{\vec{v},\mathrm{rt}}(t)\,, \tag{5.30}$$

where the last equality is valid for an equal-arm interferometer. But we just calculated these single-arm round-trip ΔT's in the previous section. Thus, the timing response of an equal-arm Michelson is simply

$$R^{ab}_{\mathrm{timing}}(f,\hat{n}) = \frac{1}{2}\left[u^a u^b\,\mathcal{T}_{\vec{u},\mathrm{rt}}(f,\hat{n}\cdot\hat{u}) - v^a v^b\,\mathcal{T}_{\vec{v},\mathrm{rt}}(f,\hat{n}\cdot\hat{v})\right]\,, \tag{5.31}$$

where we have chosen the origin of coordinates to be at the vertex of the interferometer. The phase and strain responses of a Michelson are related to the timing response by constant multiplicative factors, cf. (5.3) and (5.4), so that

$$R^{ab}_{\mathrm{phase}}(f,\hat{n}) = 2\pi\nu_0 R^{ab}_{\mathrm{timing}}(f,\hat{n})\,, \tag{5.32}$$

$$R^{ab}_{\mathrm{strain}}(f,\hat{n}) = R^{ab}_{\mathrm{timing}}(f,\hat{n})/(2L/c)\,, \tag{5.33}$$

where ν_0 is the frequency of the laser. Note that in the small-antenna limit, which is valid for the LIGO detectors below a few kHz, the strain response is given by

$$R^{ab}_{\mathrm{strain}}(f,\hat{n}) = \frac{1}{2}(u^a u^b - v^a v^b) \qquad (\text{for } fL/c \ll 1)\,. \tag{5.34}$$

Plots of the antenna patterns for the strain response to $A = +, \times$ polarized gravitational waves are given in Figure 28, for both the small-antenna limit (where we simply set $f = 0$) and at the *free-spectral range* of the interferometer, $f = f_{\mathrm{fsr}} \equiv c/(2L)$. Similar plots of the antenna patterns for unpolarized gravitational waves are given in Figure 29. In Figure 30 we show colorbar plots of the antenna patterns for the strain response to unpolarized gravitational waves for the LIGO Hanford and Virgo interferometers (located in Hanford, WA and Cascina, Italy, respectively), again evaluated in the small-antenna limit.

We can also calculate the strain response of an interferometer to the gradient and curl tensor spherical harmonic components $\{a^G_{(lm)}(f), a^C_{(lm)}(f)\}$ by performing the integration in (5.12). As shown in Appendix E of [69], this leads to

$$R^G_{(lm)}(f) = \delta_{l2}\frac{4\pi}{5}\sqrt{\frac{1}{3}}\left[Y_{2m}(\hat{u}) - Y_{2m}(\hat{v})\right]\,, \qquad R^C_{(lm)}(f) = 0\,, \tag{5.35}$$

for an interferometer in the small-antenna limit, where the vertex is at the origin of coordinates, and $\hat{u}$, $\hat{v}$ are unit vectors pointing in the direction of the interferometer arms. Similar to (5.24) for pulsar timing, the curl response is again identically zero. We will discuss the consequences of this result in more detail in Section 7.5.7, in the context of phase-coherent mapping of anisotropic gravitational-wave backgrounds.

[10] Although Figure 27 shows $\hat{u}$ and $\hat{v}$ making right angles with one another, the following calculation is valid for $\hat{u}$ and $\hat{v}$ separated by an *arbitrary* angle.

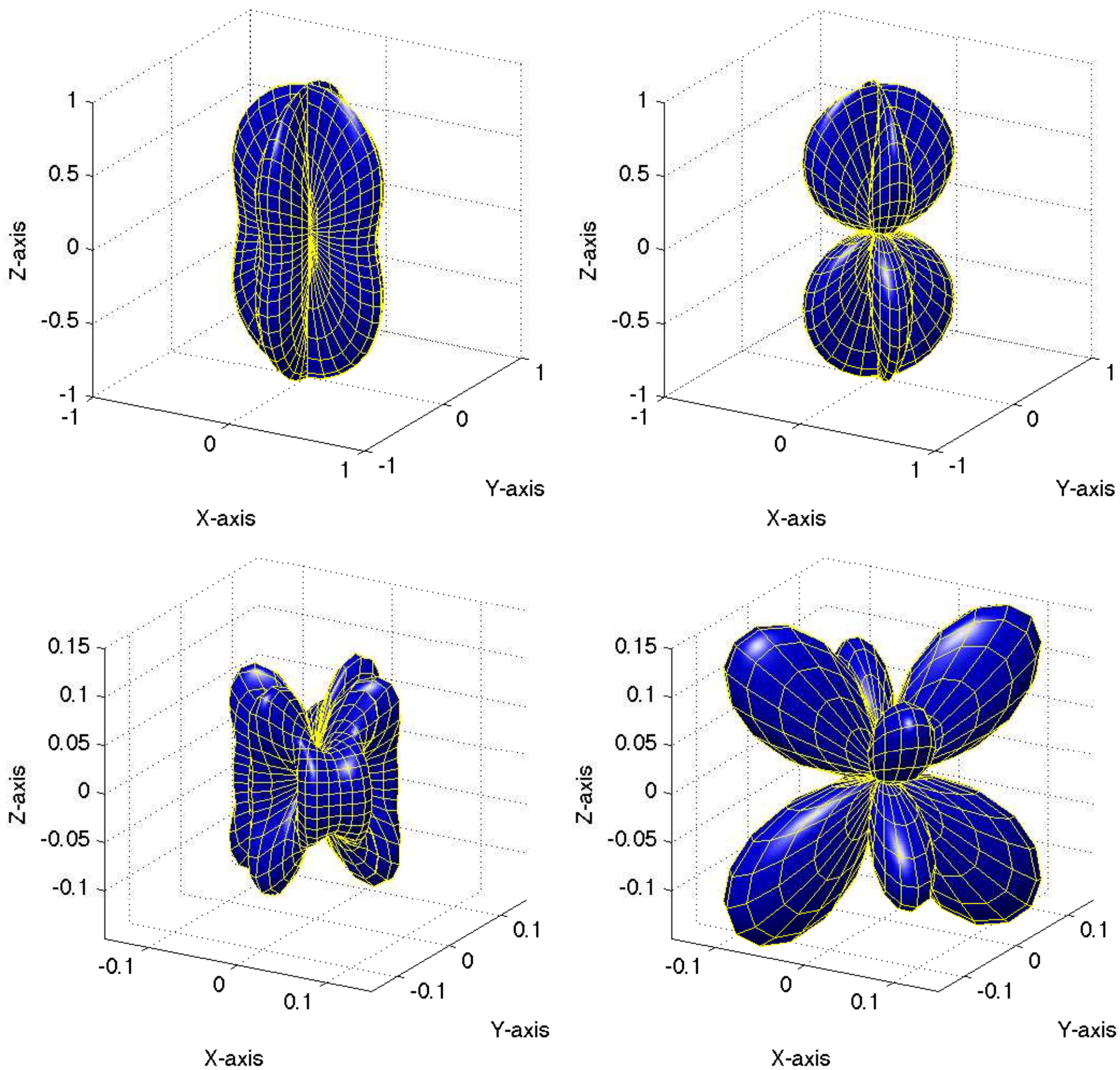


Figure 28: Antenna patterns for Michelson interferometer strain response $|R^{+}_{\mathrm{strain}}|$ and $|R^{\times}_{\mathrm{strain}}|$ evaluated in the small-antenna limit, $f = 0$ (top two plots) and at the free-spectral range frequency, $f = c/(2L)$ (bottom two plots). The interferometer arms point in the $\hat{x}$ and $\hat{y}$ directions. Note the change in the scale of the axes between the top and bottom two plots.

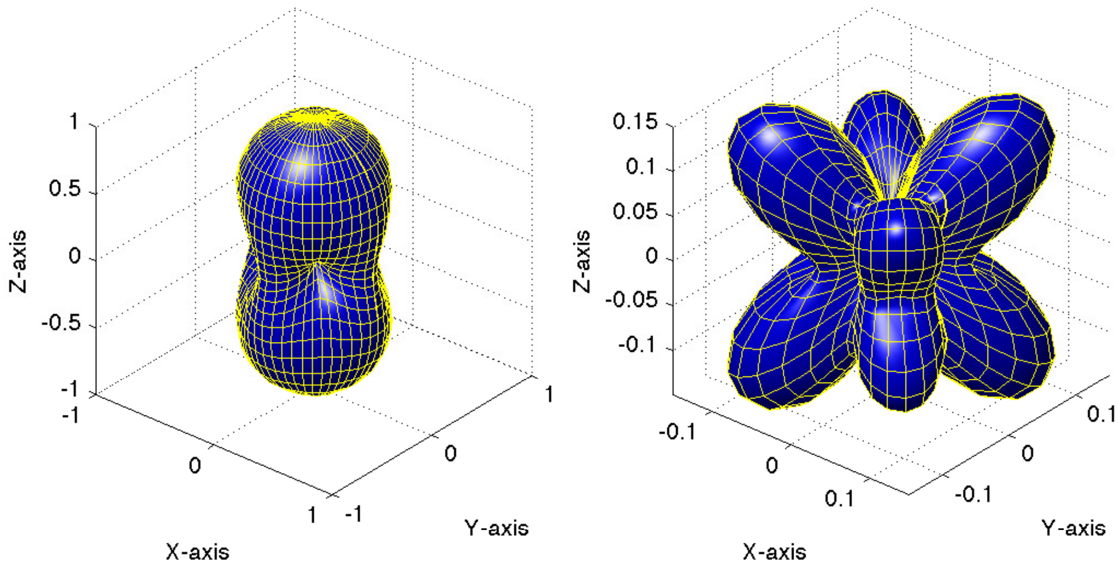


Figure 29: Antenna pattern for Michelson interferometer strain response to unpolarized gravitational waves evaluated in the small-antenna limit, $f = 0$ (left plot) and at the free-spectral range frequency, $f = c/(2L)$ (right plot). The interferometer arms point in the $\hat{x}$ and $\hat{y}$ directions. Note the change in the scale of the axes between the two plots.

5.3 Overlap functions

As mentioned in Section 4, a stochastic gravitational-wave background manifests itself as a non-vanishing correlation between the data taken by two or more detectors. This correlation differs, in general, from that due to instrumental noise, allowing us to distinguish between a stochastic gravitational-wave signal and other noise sources. In this section, we calculate the expected correlation due to a gravitational-wave background, allowing for non-trivial detector response functions and non-trivial detector geometry. Interested readers can find more details in [82, 42, 43, 66, 65].

5.3.1 Definition

Let d_I and d_J denote the data taken by two detectors labeled by I and J. In the presence of a gravitational wave, these data will have the form

$$d_I = h_I + n_I\,, \tag{5.36}$$

$$d_J = h_J + n_J\,, \tag{5.37}$$

where $h_{I,J}$ denote the response of detectors I, J to the gravitational wave, and $n_{I,J}$ denote the contribution from instrumental noise. If the instrumental noise in the two detectors are uncorrelated with one another, it follows that the expected correlation of the data is just the expected correlation of the detector responses, $\langle d_I d_J \rangle = \langle h_I h_J \rangle$. If we also assume

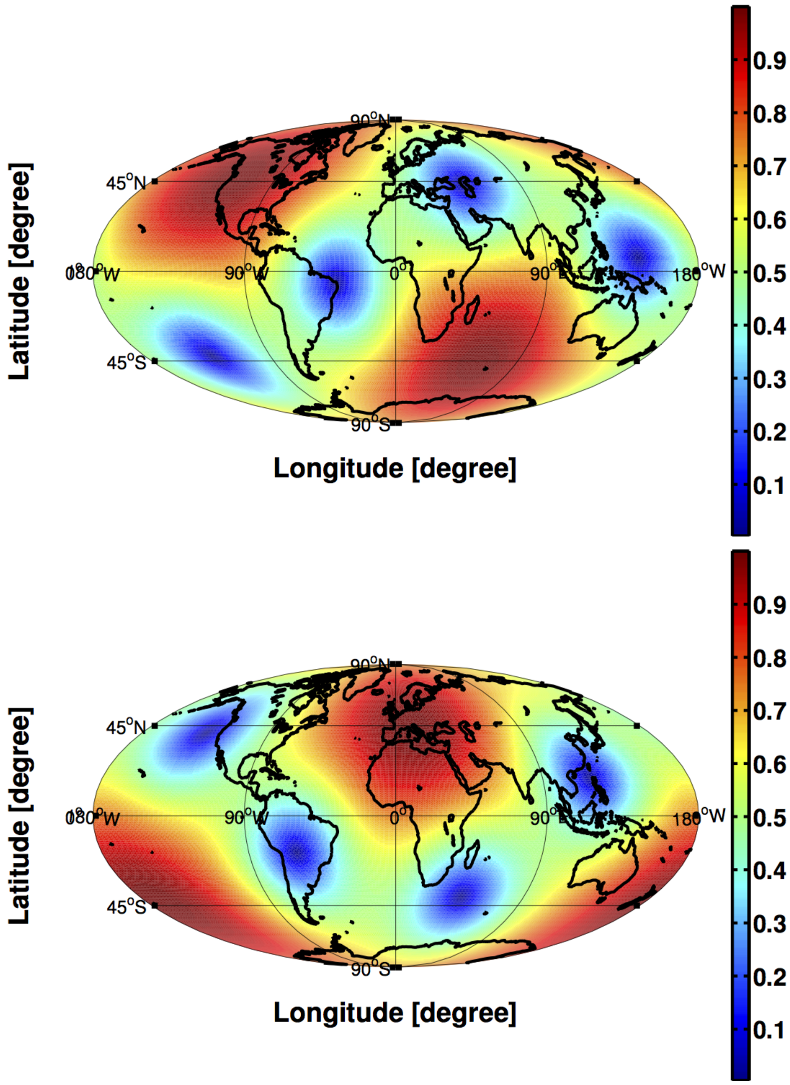


Figure 30: Antenna patterns for the strain response to unpolarized gravitational waves for the LIGO Hanford (top panel) and Virgo (bottom panel) interferometers evaluated in the small-antenna limit. The antenna patterns are represented as colorbar plots on a Mollweide projection of the Earth. Note that the maxima of the antenna patterns (the centers of the red regions) are directly above (and below) the location of the two interferometers—in Hanford, WA and Cascina, Italy, respectively. The blue regions correspond to the minima of the antenna patterns—i.e., the 'dimples' in the left panel plot of Figure 29.

that the gravitational wave is due to a stationary, Gaussian, isotropic, and unpolarized stochastic background, then

$$\langle h_I(t)h_J(t')\rangle = \frac{1}{2}\int_{-\infty}^{\infty} df\, e^{i2\pi f(t-t')}\Gamma_{IJ}(f)S_h(f)\,, \tag{5.38}$$

where where $S_h(f)$ is the one-sided strain power spectral density of the gravitational-wave background, computed from the expectation values of the Fourier components of the metric perturbations (2.13), and

$$\Gamma_{IJ}(f) \equiv \frac{1}{8\pi}\int d^2\Omega_{\hat{n}} \sum_A R_I^A(f,\hat{n})R_J^{A*}(f,\hat{n}) \tag{5.39}$$

is the so-called *overlap function* for the two detectors I, J written in terms of the polarization-basis response function $R^A_{I,J}(f,\hat{n})$.[11] In terms of the tensor spherical harmonic-basis response functions $R^P_{I,J(lm)}(f)$, we would have

$$\Gamma_{IJ}(f) = \frac{1}{8\pi}\sum_{(lm)}\sum_P R^P_{I(lm)}(f)R^{P*}_{J(lm)}(f)\,, \tag{5.40}$$

where $P = \{G, C\}$ for the gradient and curl tensor spherical harmonic components.

5.3.2 Interpretation

The overlap function $\Gamma_{IJ}(f)$ quantifies the reduction in sensitivity of the cross-correlation to a stochastic gravitational-wave background due to the non-trivial response of the detectors and their separation and orientation relative to one another. This meaning of the overlap function is most easily seen in the frequency domain, where (5.38) becomes

$$\langle \tilde{h}_I(f)\tilde{h}_J^*(f')\rangle = \frac{1}{2}\delta(f-f')\,\Gamma_{IJ}(f)S_h(f)\,. \tag{5.41}$$

This implies

$$\tilde{C}_{h_Ih_J}(f) = \Gamma_{IJ}(f)S_h(f)\,, \tag{5.42}$$

where $\tilde{C}_{h_Ih_J}(f)$ is the (one-sided) cross-spectrum of the response in the two detectors. Thus, $\Gamma_{IJ}(f)$ can be interpreted as the transfer function between gravitational-wave strain power $S_h(f)$ and detector response cross-power $\tilde{C}_{h_Ih_J}(f)$.

Expression (5.39) for the overlap function involves four length scales: the lengths of the two detectors, L_I and L_J, which appear in the response functions $R^A_{I,J}(f,\hat{n})$; the separation of the detectors, $s \equiv |\vec{x}_I - \vec{x}_J|$, which appears in the exponential factor; and

[11] Recall from Footnote 8 that the phase factors $e^{i2\pi f\hat{n}\cdot\vec{x}_{I,J}/c}$ are already contained in our definition of the response functions $R^A_{I,J}(f,k)$. If we explicitly display this dependence then

$$\Gamma_{IJ}(f) \equiv \frac{1}{8\pi}\int d^2\Omega_{\hat{n}}\sum_A \bar{R}_I^A(f,\hat{n})\bar{R}_J^{A*}(f,\hat{n})e^{i2\pi f\hat{n}\cdot(\vec{x}_I-\vec{x}_J)/c}\,,$$

where $\bar{R}^A_{I,J}(f,\hat{n}) \equiv \bar{R}^{ab}_{I,J}(f,\hat{n})e^A_{ab}(\hat{n})$. One often sees this latter expression for $\Gamma_{IJ}(f)$ in the literature.

the wavelength of the gravitational waves, $\lambda = c/f$. In general, one has to evaluate the integral in (5.39) *numerically*, due to the non-trivial frequency dependence of the response functions. However, as we shall see in Section 5.4, in certain limiting cases of the ratio of these length scales, we can do the integral *analytically* and obtain relatively simple expressions for the overlap function in terms of spherical Bessel or trigonometric functions. This is the case for ground-based interferometers, which operate in the *small-antenna limit*—i.e., $fL/c \ll 1$ for both detectors, even though the separation can be large compared to the wavelength, $fs/c \gtrsim 1$. It is also the case for pulsar timing arrays, which operate in the *large-antenna, small-separation limit*, since $fL/c \gg 1$ for each pulsar and $fs/c \ll 1$ for different radio receivers on Earth. (The Earth effectively resides at the solar system barycenter relative to the wavelength of the gravitational waves relevant for pulsar timing.)

5.3.3 Normalization

It is often convenient to define a *normalized* overlap function $\gamma_{IJ}(f) \propto \Gamma_{IJ}(f)$ by requiring that $\gamma_{IJ}(0) = 1$ for two detectors that are co-located and co-aligned. For the strain response of two identical equal-arm Michelson interferometers, this leads to the relation

$$\gamma_{IJ}(f) = \frac{5}{\sin^2\beta}\,\Gamma_{IJ}(f) \tag{5.43}$$

where β is the opening angle between the two arms ($\pi/2$ for LIGO and $\pi/3$ for LISA).

5.3.4 Auto-correlated response

To obtain the *auto-correlated* response of a *single* detector, we can simply set $I = J$ in the previous expressions. This means that the gravitational-wave strain power $S_h(f)$ and the detector response power $P_{h_I}(f)$ in detector I are related by

$$P_{h_I}(f) = \Gamma_{II}(f) S_h(f)\,, \tag{5.44}$$

where

$$\Gamma_{II}(f) = \frac{1}{8\pi}\int d^2\Omega_{\hat{n}} \sum_A |R^A_I(f,\hat{n})|^2\,. \tag{5.45}$$

Note that $\Gamma_{II}(f)$ is just the square of the antenna pattern for the response to unpolarized gravitational waves *integrated over the whole sky.* A plot of the normalized transfer function $\gamma_{II}(f)$ for the strain response of an equal-arm Michelson interferometer is shown in Figure 31. Compared to Figure 24 for the timing transfer function $|\mathcal{T}_{\vec{u}}(f,0)|$ for one-way photon propagation evaluated at normal incidence of the gravitational wave, we see that the relevant frequency scale for an equal-arm Michelson is $c/(2L)$ (as opposed to c/L) due to the round-trip motion of the photons. Also, the hard nulls in Figure 24 have been softened into *dips* due to averaging of the waves over the whole sky. The high-frequency 'bumps' for $\gamma_{II}(f)$ are lower than those for $|\mathcal{T}_{\vec{u}}(f,0)|$ due to the squaring of $|R^A_I(f,\hat{n})|$ which enters into the defintion of $\Gamma_{II}(f)$ (and $\gamma_{II}(f)$). Figure 32 is an extended version of Figure 31, with the appropriate frequency ranges for ground-based interferometers (like LIGO), space-based interferometers (like LISA), spacecraft Doppler tracking, and pulsar timing searches indicated on the plot. See also Table 5 for more details.

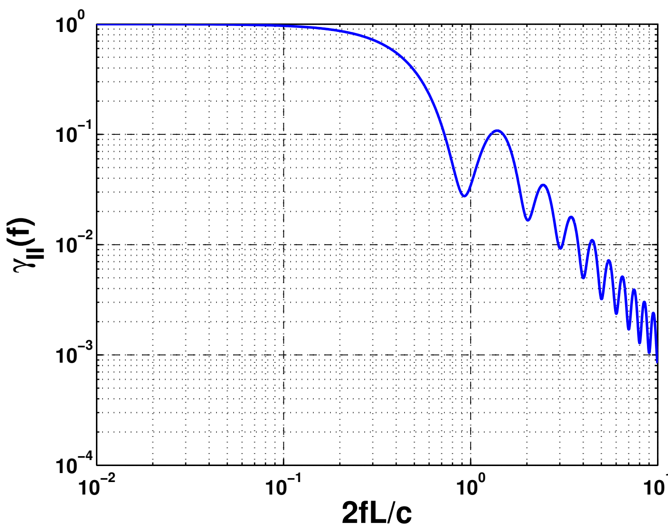


Figure 31: A plot of the normalized transfer function $\gamma_{II}(f)$ for the strain response of an equal-arm Michelson interferometer. The dips in the transfer function occur around integer multiples of $c/(2L)$.

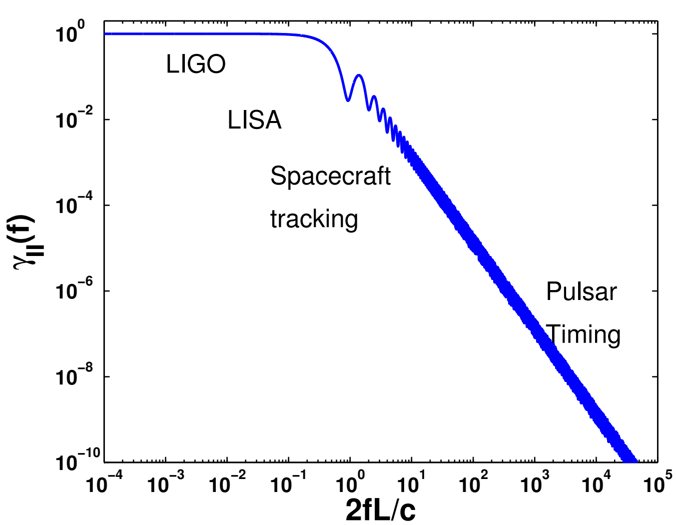


Figure 32: An extension of Figure 31 to lower and higher frequencies, and plotted on a log-log scale. The position of the labels show the relative location of the frequency bands for gravitational-wave searches using ground-based interferometers like LIGO, space-based interferometers like LISA, spacecraft Doppler tracking, and pulsar timing arrays, expressed in units of $c/(2L)$. See also Table 5 for more details.

5.4 Examples of overlap functions

5.4.1 LHO-LLO overlap function

As mentioned above, Earth-based interferometers like LIGO operate in the small-antenna limit where $fL/c \ll 1$. This implies that the associated response functions are well-approximated by the expression in (5.34). If we denote the unit vectors along the two arms of one Earth-based interferometer by $\hat{u}_1$ and $\hat{v}_1$, and the corresponding unit vectors of a second Earth-based interferometer by $\hat{u}_2$ and $\hat{v}_2$, then the strain responses in the two interferometers are simply

$$\begin{aligned} R^A_{1,\text{strain}}(f,\hat{n}) &\simeq D_1^{ab}\, e^A_{ab}(\hat{n}) e^{i2\pi f\hat{n}\cdot\vec{x}_1/c}\,, \\ R^A_{2,\text{strain}}(f,\hat{n}) &\simeq D_2^{ab}\, e^A_{ab}(\hat{n}) e^{i2\pi f\hat{n}\cdot\vec{x}_2/c}\,, \end{aligned} \tag{5.46}$$

where

$$D_1^{ab} \equiv \frac{1}{2}\left(u_1^a u_1^b - v_1^a v_1^b\right)\,, \qquad D_2^{ab} \equiv \frac{1}{2}\left(u_2^a u_2^b - v_2^a v_2^b\right)\,, \tag{5.47}$$

and $\vec{x}_1$ and $\vec{x}_2$ denote the vertices of the two interferometers. The tensors D_1^{ab}, D_2^{ab} defined above are called *detector tensors*; they are symmetric and trace-free with respect to their ab indices. In terms of the detector tensors, the overlap function becomes

$$\Gamma_{12}(f) = D_1^{ab} D_2^{cd} \Gamma_{abcd}(\Delta\vec{x})\,, \tag{5.48}$$

where

$$\Gamma_{abcd}(\Delta\vec{x}) \equiv \int d^2\Omega_{\hat{n}} \sum_A e^A_{ab}(\hat{n}) e^A_{cd}(\hat{n})\, e^{-i2\pi f\hat{n}\cdot\Delta\vec{x}/c} \tag{5.49}$$

and $\Delta\vec{x} \equiv \vec{x}_2 - \vec{x}_1$ is the separation vector connecting the two vertices. Thus, in the small-antenna limit, the orientation-dependence of the overlap function $\Gamma_{12}(f)$ is encoded in the detector tensors D_1^{ab}, D_2^{ab}, while the separation-dependence is encoded in $\Gamma_{abcd}(\Delta\vec{x})$. Note that Γ_{abcd} is a tensor which is symmetric under the interchanges $a \leftrightarrow b$, $c \leftrightarrow d$, and $ab \leftrightarrow cd$; it is also trace-free with respect to the ab and cd index pairs. Using these symmetry properties, one can *analytically* do the integrals in (5.49) (see e.g., [66, 26]), leading to

$$\Gamma_{12}(f) = \rho_1(\alpha)\, D_1^{ab} D_{2ab} + \rho_2(\alpha)\, D_1^{ab} D_{2a}{}^c s_b s_c + \rho_3(\alpha)\, D_1^{ab} D_2^{cd} s_a s_b s_c s_d\,, \tag{5.50}$$

where

$$\begin{bmatrix} \rho_1(\alpha) \\ \rho_2(\alpha) \\ \rho_3(\alpha) \end{bmatrix} = \frac{4\pi}{5} \begin{bmatrix} 10 & -20/\alpha & 10/\alpha^2 \\ -20 & 80/\alpha & -100/\alpha^2 \\ 5 & -50/\alpha & 175/\alpha^2 \end{bmatrix} \begin{bmatrix} j_0(\alpha) \\ j_1(\alpha) \\ j_2(\alpha) \end{bmatrix}\,, \tag{5.51}$$

and

$$\alpha \equiv 2\pi f s/c\,, \quad s \equiv |\Delta\vec{x}|\,, \quad \hat{s} \equiv \Delta\vec{x}/s\,, \tag{5.52}$$

and $j_0(\alpha)$, $j_1(\alpha)$, and $j_2(\alpha)$ are the standard spherical Bessel functions [14]. Thus, in the small-antenna limit, the overlap function for the strain response of two equal-arm Michelson interferometers can be written as a sum of the first three spherical Bessel functions with coefficients that depend on the product of the frequency and separation of the two detectors. (This analytic expression for the overlap function can also be derived using (5.40), which involves the tensor spherical harmonic repsonse functions. A detailed derivation using these response functions is given in [142].)

Figure 33 is a plot of the normalized overlap function for the strain response of the 4-km LIGO interferometers in Hanford, WA and Livingston, LA. There are several things to note about the plot: (i) The overlap function is negative as $f \to 0$. This is because the arms of the Hanford and Livingston interferometers are rotated by 90° with respect to one another. (ii) The magnitude of the overlap function at $f = 0$ is less than unity—i.e., $|\gamma_{HL}(0)| = 0.89$, even though the overlap function was normalized. This is because the planes of the Hanford and Livingston interferometers are not identical; these two detectors are separated by 27.2° as seen from the center of the Earth. (iii) The first zero of the overlap function occurs just above 60 Hz. This is roughly equal to $c/(2s) = 50$ Hz, where $s = 3000$ km is the separation between the two interferometers. Note that $f = c/(2s)$ is the frequency of a gravitational wave that has a wavelength equal to twice the separation of the two sites. For lower frequencies, the two interferometers will be driven (on average) by the same positive (or negative) part of the incident gravitational wave. For slightly higher frequencies, one interferometer will be driven by the positive (or negative) part of the incident wave, while the other interferometer will be driven by the negative (or positive) part. The zeros of the overlap function correspond to the transitions between the in-phase and out-of-phase excitations of the two interferometers.

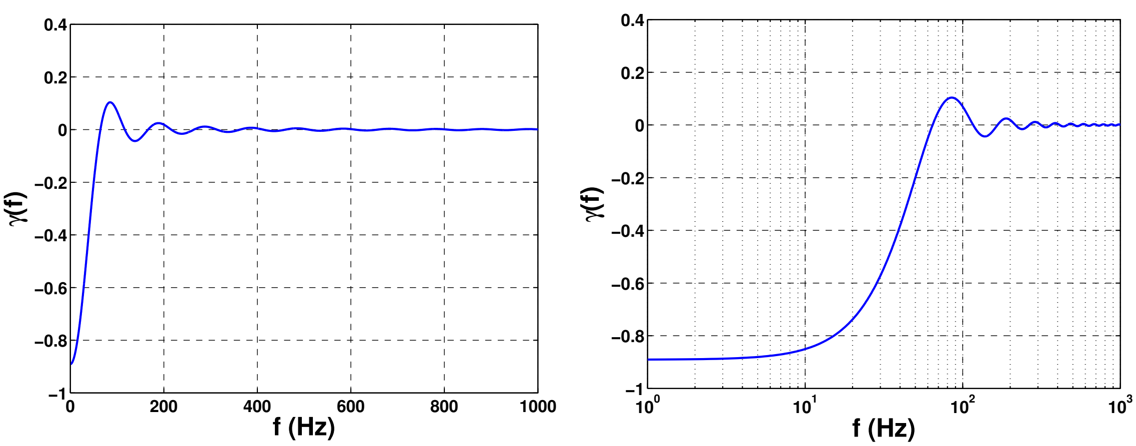


Figure 33: Overlap function for the LIGO Hanford-LIGO Livinston cross-correlation in the small-antenna limit. Left panel: linear frequency scale. Right panel: logarithmic frequency scale.

5.4.2 Big-Bang Observer overlap function

As a second example, we consider the overlap function between two LISA-like constellations oriented in a hexagram (i.e., ‘six-pointed star’) configuration as shown in Figure 34. This is one of the configurations being considered for the Big-Bang Observer (BBO), which is a proposed space mission designed to detect or put stringent limits on a cosmologically-generated gravitational-wave background [134]. The arm lengths of the two interferometers, with vertices $\vec{x}_1$ and $\vec{x}_2$, are taken to be $L = 5 \times 10^6$ km. The opening angle for the two interferometers is $\beta = 60°$. For this example, we calculate the normalized overlap

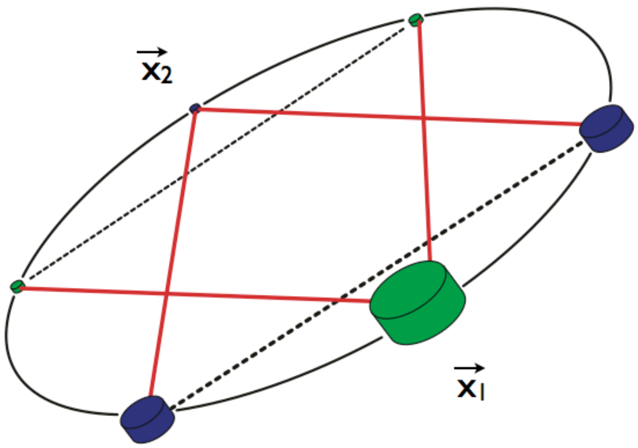


Figure 34: Hexagram configuration for the cross-correlation of two LISA-like detectors, relevant for the proposed Big-Bang Observer space mission. Spacecraft, which house lasers and freely-falling test masses, are located at each vertex of the hexagram. The vectors $\vec{x}_1$ and $\vec{x}_2$ denote the vertices of two equal-arm Michelson interferometers, with opening angle $\beta = 60°$. (Figure taken from [47].)

function for strain response numerically, since the small-antenna limit is not valid for the high-frequency end of the sensitivity band. A plot of the normalized overlap function is given in Figure 35.

5.4.3 Pulsar timing overlap function (Hellings and Downs curve)

As our final example, we consider the overlap function for timing residual measurements from an array of N pulsars, labeled by index $I = 1, 2, \cdots, N$. Each pulsar defines a one-way tracking beam detector with the position of pulsar I at $\vec{p}_I$ and the postion of detector I (i.e., a radio receiver on Earth) by $\vec{x}_I$. For convenience, we will take the origin of coordinates to lie at the solar system barycenter. Since the diameter of the Earth ($\sim 10^4$ km) and its distance from the Sun ($\sim 10^8$ km) are both small compared to the wavelength of gravitational waves relevant for pulsar timing ($\lambda = c/f \sim 10^{13}$ km), we can effectively set $\vec{x}_I \approx \vec{x}_J \approx \vec{0}$ in the argument of the exponential term that enters expression (5.39) for the overlap function. Thus,

$$\Gamma_{IJ}(f) = \frac{1}{(2\pi f)^2} \int d^2\Omega_{\hat{n}} \sum_A \frac{1}{2} u_I^a u_I^b e_{ab}^A(\hat{n}) \, \frac{1}{2} u_J^c u_J^d e_{cd}^A(\hat{n}) e^{i2\pi f \hat{n}\cdot\cancel{(\vec{x}_I - \vec{x}_J)}/c} \times \frac{1}{1+\hat{n}\cdot\hat{u}_I} \frac{1}{1+\hat{n}\cdot\hat{u}_J} \left[1 - \cancel{e^{-i\frac{2\pi f L_I}{c}(1+\hat{n}\cdot\hat{u}_I)}}\right] \left[1 - \cancel{e^{+i\frac{2\pi f L_J}{c}(1+\hat{n}\cdot\hat{u}_J)}}\right] \tag{5.53}$$

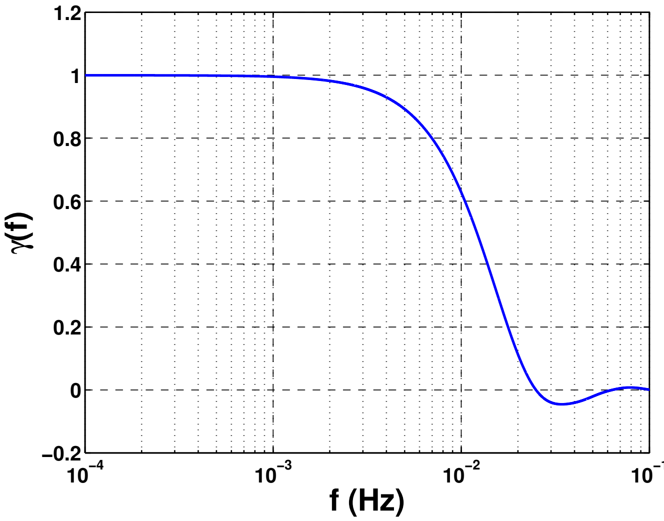


Figure 35: Plot of the normalized overlap function for strain response for the hexagram configuration shown in Figure 34.

where the unit vectors $\hat{u}_I$, $\hat{u}_J$ are defined by $\vec{x}_I = \vec{p}_I + L_I\hat{u}_I$, where L_I is the distance to pulsar I. But since $\vec{x}_I \approx \vec{0}$, it follows that $\hat{u}_I$ and $\hat{u}_J$ are just unit vectors pointing from the location of pulsars I and J *toward* the SSB. For distinct pulsars ($I \neq J$), we can ignore the exponential terms in the square brackets, since $fL/c \gg 1$ for $L \sim 1$ kpc ($= 3 \times 10^{16}$ km) implies that $e^{-i2\pi f L_I(1+\hat{n}\cdot\hat{u}_I)/c}$ and its product with the corresponding term for pulsar J are rapidly varying functions of $\hat{n}$ and do not contribute significantly when integrated over the whole sky [82, 28]. (For a single pulsar ($I = J$), the product of the two exponential terms equals 1 and hence cannot be ignored.) With these simplifications, the integral can be done analytically [82, 28, 97]. The result is

$$\Gamma_{IJ}(f) = \frac{1}{(2\pi f)^2}\,\frac{1}{3}\,\chi(\zeta_{IJ})\,, \tag{5.54}$$

where

$$\chi(\zeta_{IJ}) \equiv \frac{3}{2}\left(\frac{1-\cos\zeta_{IJ}}{2}\right)\ln\left(\frac{1-\cos\zeta_{IJ}}{2}\right) - \frac{1}{4}\left(\frac{1-\cos\zeta_{IJ}}{2}\right) + \frac{1}{2} + \frac{1}{2}\delta_{IJ}\,, \tag{5.55}$$

and ζ_{IJ} is the angle between the two pulsars I and J relative to the solar system barycenter. (For Doppler frequency measurments, the overlap function is *independent* of frequency, $\Gamma_{IJ} = \chi(\zeta_{IJ})/3$.) $\chi(\zeta)$ is the *Hellings and Downs* function [82]; it depends only on the angular separation of a pair of pulsars. The normalization was chosen so that for a single pulsar, $\chi(0) = 1$ (for two *distinct* pulsars occupying the same angular position on the sky, $\chi(0) = 0.5$). A plot of the Hellings and Downs curve is given in Figure 36.

A couple of remarks are in order: (i) The Hellings and Downs curve is *independent* of frequency; it is a function of the *angle* ζ between different pulsar pairs. This contrasts with

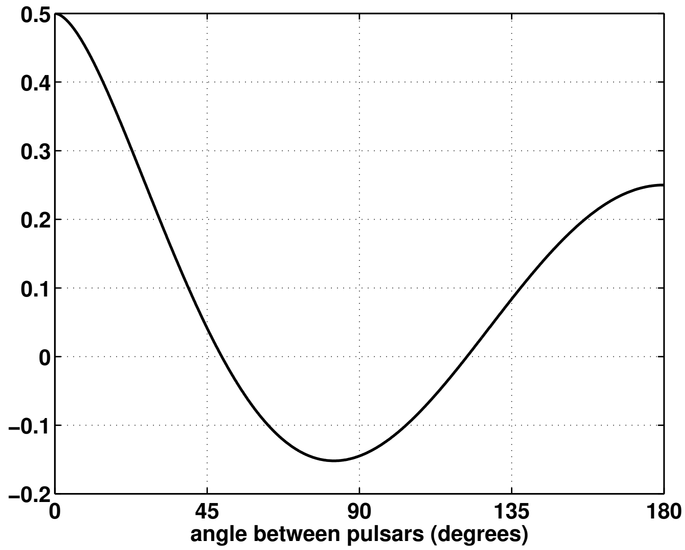


Figure 36: Plot of the Hellings and Downs curve as a function of the angular separation between two distinct pulsars.

the overlap functions for the two LIGO interferometers and for BBO given in Figures 33 and 35. These overlap functions were calculated for a fixed pair of detectors; they are functions instead of the *frequency* of the gravitational wave. (ii) The value of the Hellings and Downs function $\chi(\zeta_{IJ})$ for a pair of pulsars I, J can be written as a Legendre series in the cosine of the angle between the two pulsars. This follows immediately if one uses (5.40) for the overlap function and (5.24) for the pulsar timing response functions in the tensor spherical harmonic basis. As shown in [69]:

$$\chi(\zeta_{IJ}) = \frac{3}{4}\sum_{l=2}^{\infty}({}^{(2)}N_l)^2(2l+1)P_l(\hat{p}_I\cdot\hat{p}_j)\,, \tag{5.56}$$

where $\hat{p}_I$ and $\hat{p}_J$ are unit vectors that point in the directions to the two pulsars. A Legendre series expansion out to $l_{\max} = 4$ (i.e., only three terms) gives very good agreement with the exact expression for the Hellings and Downs function, except for very small angular separations. This is illustrated in Figure 37.

5.5 Moving detectors

So far, we have ignored any time-dependence in the detector response introduced by the motion of the detectors relative to the gravitational-wave source. In general, this relative motion produces a *modulation* in both the *amplitude* and the *phase* of the response of a detector to a monochromatic, plane-fronted gravitational wave [54]. For Earth-based interferometers like LIGO, the modulation is due to both the Earth's daily rotation and yearly orbital motion around the Sun. For space-based interferometers like LISA, the

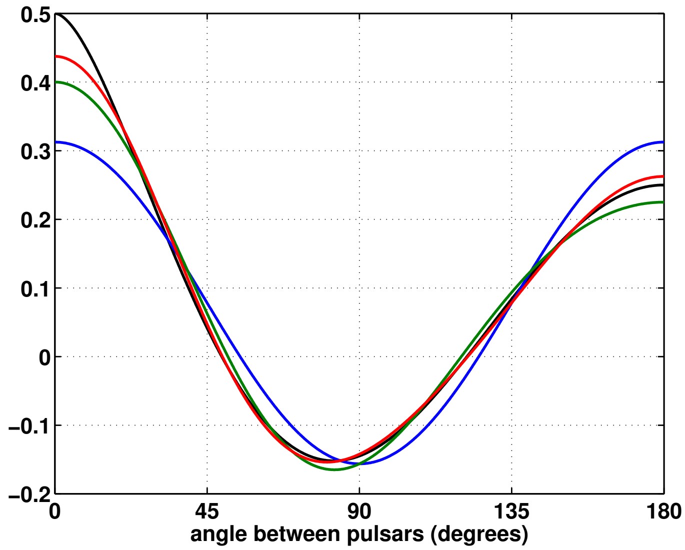


Figure 37: Comparison of the exact expression of the Hellings and Downs curve (black) with Legendre series approximations for different values of $l_{\max}$. The blue, green, and red curves correspond to $l_{\max} = 2$, 3, and 4, respectively.

modulation is due to the motion of the individual spacecraft as they orbit the Sun with a period of one year. For example, for the original LISA design, three spacecraft fly in an equilateral-triangle configuration around the Sun. The center-of-mass (or guiding center) of the configuration moves in a circular orbit of radius 1 AU, at an angle of 20° behind Earth, while the configuration 'cartwheels' in retrograde motion about the guiding center, also with a period of one year (see Figure 38).

5.5.1 Mononchromatic plane waves

The phase modulation of a monochromatic plane wave will have contributions from both the time-varying *orientation* of the detector as well as the detector's *translational* motion relative the source. The time-varying orientation leads to changes in the response of the detector to the + and × polarization components of the wave, $|R^+h_+|$ and $|R^\times h_\times|$. The translational motion leads to a Doppler shift in the observed frequency of the wave, which is proportional to v/c times the nominal frequency, where v is velocity of the detector relative to the source:

$$\Delta_D f = \frac{1}{2\pi}\frac{d\varphi_D(t)}{dt} = -f\hat{n}\cdot\vec{v}(t)/c\,. \tag{5.57}$$

For example, for a monochromatic source with $f = 100$ Hz observed by ground-based detectors like LIGO, the Earth's daily rotational motion ($v \approx 500$ m/s) produces a Doppler shift of order $\sim 10^{-4}$ Hz, while the Earth's yearly orbital motion ($v \approx 3\times10^4$ m/s), produces a shift of order $\sim 10^{-2}$ Hz. A matched filter search for a sinusoid must take

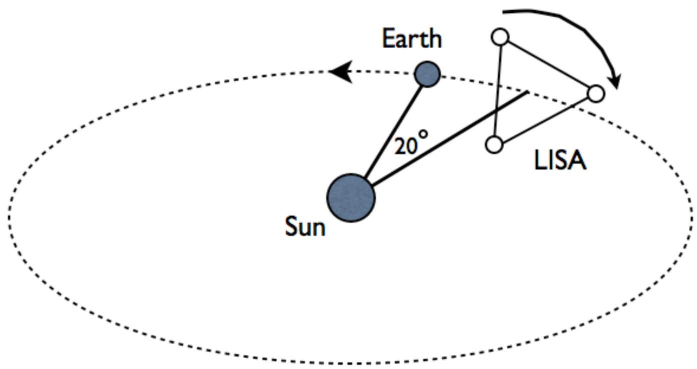


Figure 38: Original LISA configuration: The center-of-mass of the equilateral-triangle configuration of spacecraft orbits the Sun in a circle of radius 1 AU, 20° behind Earth, while the configuration 'cartwheels' in retrograde motion about the center-of-mass, also with a period of one year. (Figure adapted from [47].)

this latter modulation into account, as the frequency shift is larger than the width of a frequency bin for a typical search for such a signal.

5.5.2 Stochastic backgrounds

For stochastic gravitational-wave backgrounds, things are slightly more complicated as the signal is an incoherent sum of sinusoidal plane waves having different amplitudes, frequencies, and phases, and coming from different directions on the sky (2.1). But since the signal is *broad-band*, the Doppler shift associated with the phase modulation of the individual component plane waves is not important, as the gravitational-wave signal power is (at worst) shuffled into nearby bins.[12] On the other hand, the amplitude modulation of the signal, due to the time-varying orientation of a detector, *can* be significant if the background is *anisotropic*—i.e., stronger coming from certain directions on the sky than from others. (We will discuss searches for anisotropic backgrounds in detail in Section 7.) As the lobes of the antenna pattern sweep through the "hot" and "cold" spots of the anisotropic background, the amplitude of the signal is modulated in time.

Figure 39 shows the expected time-domain output of a particular Michelson combination, $X(t)$, of the LISA data over a two-year period. The combined signal (red) consists of both detector noise (black) and the confusion-limited gravitational-wave signal from the galactic population of compact white-dwarf binaries. At freuquencies $\sim 10^{-4} - 10^{-3}$ Hz, which corresponds to the lower end of LISA's sensitivity band, the contribution from these binaries dominates the detector noise. The modulation of the detector output is clearly visible in the figure. The peaks in amplitude are more than 50% larger than the minim-

[12] Actually, the bin size for a typical LIGO search for a stochastic background is *larger* than the $\sim 10^{-2}$ Hz Doppler shift due to the Earth's orbital motion around the Sun.

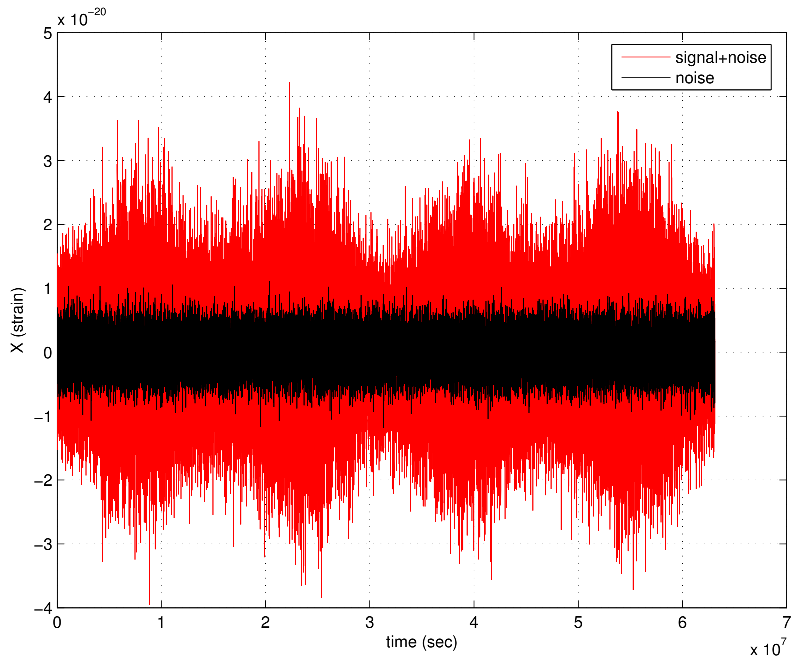


Figure 39: The time-domain output of a particular Michelson combination, $X(t)$, of the LISA data over a two-year period. The contribution from the detector noise is shown in black. The combined output, consisting of both detector noise and the confusion noise from the Galactic population of compact white-dwarf binaries, is shown in red. The modulation in the amplitude is due to the time-varying orientation of the LISA constellation as it performs a 'cart-wheel' in its 1-year orbit around the Sun (Figure 38). The amplitude of the output is largest when the main lobes of LISA's antenna pattern points in the general direction of the galactic center. (Data provided by Matt Benacquista.)

ima; they repeat on a 6 month time scale, as expected from LISA's yearly orbital motion around the Sun (Figure 38).

Figure 40 is a single frame of an animation showing the time evolution of the LISA antenna pattern, represented as as a colorbar plot on a Mollweide projection of the sky in ecliptic coordinates. The peaks in the detector output that we saw earlier in Figure 39 correspond to those times when the maxima of the antenna pattern point in the general direction of the galactic center, $(\mathrm{lon}, \mathrm{lat}) = (-93.3°, -5.6°)$ in ecliptic coordinates.[13] The motion of the LISA constellation was taken from [54], and the antenna pattern was calculated for the X-Michelson combination of the LISA data, assuming the small-antenna approximation for the interferometer response functions. The full animation corresponds to LISA's orbital period of 1 year. Go to `http://www.livingreviews.org/` to view the animation.

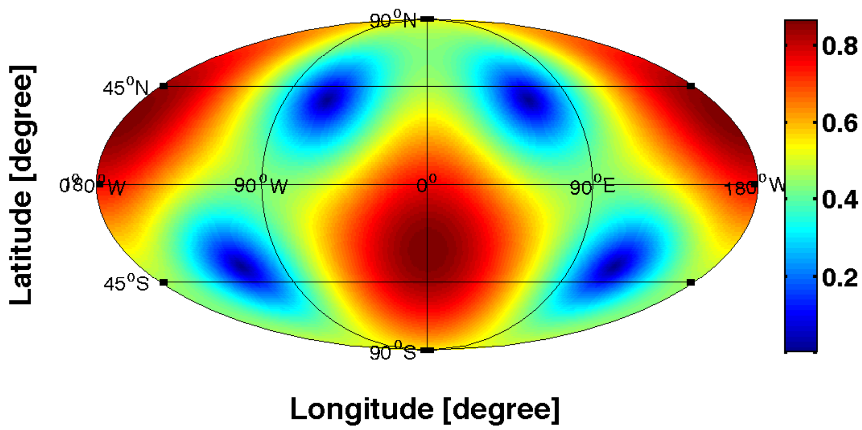


Figure 40: A single frame of an animation showing the time evolution of the LISA antenna pattern, represented as a colorbar plot on a Mollweide projection of the sky in ecliptic coordinates. Maxima (minima) of the antenna pattern are shown by the red (blue) regions. The full animation corresponds to a period of 1 year. To view the animation, please go to the online version of this review article at `http://www.livingreviews.org/`.

5.5.3 Rotational and orbital motion of Earth-based detectors

As mentioned above, given the broad-band nature of a stochastic signal, the Doppler shift associated with the motion of a detector does not play an important role for stochastic background searches. This means that we can effectively ignore the velocity of a detector, and treat its motion in as *quasi-static*. So, for example, the motion of a single Earth-based detector like LIGO can be thought of as synthesizing a *set* of *static* virtual detectors located along a quasicircular ring 1 AU from the solar system barycenter [142]. Each

[13] In equatorial coordinates, the galactic center is located at $(\mathrm{ra}, \mathrm{dec}) = (-6^{\mathrm{h}}15^{\mathrm{m}}, -29°)$.

virtual detector in this set observes the gravitational-wave background from a different spatial location and with a different orientation.

As described in [142], the relevant time-scale for a set of virtual detectors is the time over which measurements made by the different virtual detectors are *correlated* with one another. Basically, we want two neighboring virtual detectors to be spaced far enough apart that they provide *independent* information about the background. For a gravitational wave of frequency f, the minimal separation corresponds to $|\Delta\vec{x}| \approx \lambda/2$, where $\lambda = c/f$ is the wavelength of the gravitational wave. For smaller separations, the two detectors will be driven in coincidence (on average), as discussed in item (iii) in the text preceding Figure 33. Writing $|\Delta\vec{x}| = v\Delta t$ and solving for Δt yields

$$\Delta t \approx \frac{\lambda}{2v} = \frac{c}{2vf} \equiv t_{\rm corr}\,, \tag{5.58}$$

where $t_{\rm corr}$ is the correlation time-scale. For $\Delta t \lesssim t_{\rm corr}$, the measurements taken by the two virtual detectors will be correlated with one another; for $\Delta t \gtrsim t_{\rm corr}$ the measurements will be uncorrelated with one another.

As a concrete example, let us consider a gravitational wave having frequency $f = 100$ Hz, and calculate the correlation time scale for the Earth's rotational and orbital motion, treated independently. Since $v \approx 500$ m/s for daily rotation and $v \approx 3\times 10^4$ m/s for orbital motion, we get

$$\begin{aligned} &t_{\rm corr} \approx 3000\ {\rm s} \quad ({\rm rotational\ motion})\,,\\ &t_{\rm corr} \approx 50\ {\rm s} \qquad ({\rm orbital\ motion})\,. \end{aligned} \tag{5.59}$$

Thus, the orbital motion of the Earth around the Sun will more rapidly synthesize a large network of independent detectors from the motion of a single detector, compared to just rotational motion.

We can confirm these approximate results by plotting the overlap function at $f = 100$ Hz for two virtual interferometers synthesized by the Earth's rotational and orbital motion as function of time. This is done in Figure 41, assuming an isotropic and unpolarized stochastic background, and using the small-antenna approximation to calculate the detector response functions. The left-hand plot is for a set of virtual interferometers synthesized by the daily rotation of a detector located on the Earth's equator, with no orbital motion. The center of the Earth is fixed at the solar system barycenter, and the virtual interferometers have one arm pointing North and the other pointing East. One sees from the plot that the virtual interferometers decorrelate on a timescale of roughly an hour, consistent with (5.59), and recorrelate after 24 hrs when the original detector returns to its starting position. The right-hand plot is for a set of virtual interferometers at 1 AU from the solar system barycenter, associated with Earth's yearly orbital motion. There is no rotational motion for this case, as the interferometers are located at the center of the Earth in its orbit around the Sun, with the orientation of the interferometer arms unchanged by the orbital motion. Here we see that the virtual interferometers decorrelate on a timescale of roughly 1 minute, again consistent with (5.59). They will recorrelate only after 1 year (not shown on the plot). Since the orbital velocity of the Earth is much larger than the velocity of a detector on the surface of the Earth due to the Earth's daily rotational motion, the virtual interferometers associated with orbital motion build up a larger separation and decorrelate on a much shorter time scale.

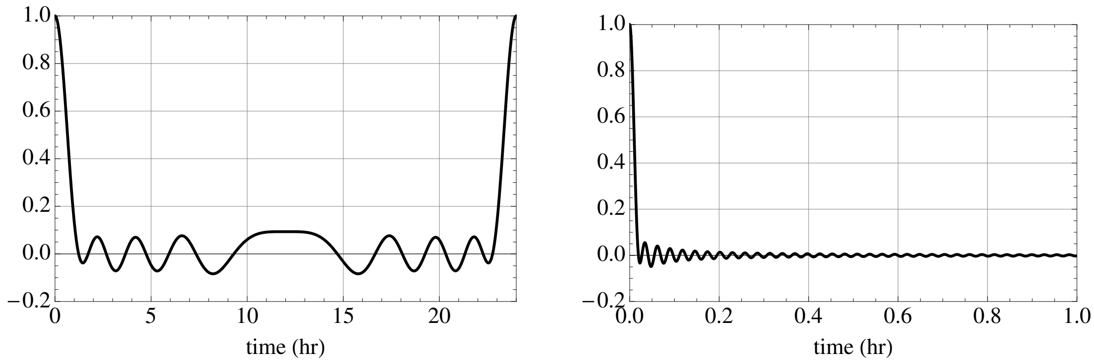


Figure 41: Overlap function at $f = 100$ Hz for two virtual interferometers as a function of time. The left-hand plot is for a set of virtual interferometers located on Earth's equator, associated with Earth's daily rotational motion. The right-hand plot is for a set of virtual interferometers at 1 AU from the SSB, associated with Earth's yearly orbital motion. The first zero-crossing times in these two plots are consistent with the correlation times given in (5.59). Figure taken from [142].

We will return to this idea of using the motion of a detector to synthesize a set of static virtual detectors when we discuss a *phase-coherent* approach for mapping anisotropic gravitational-wave backgrounds in Section 7.5.

6 Optimal filtering

Filters are for cigarrettes and coffee. *Cassandra Clare*

Optimal filtering, in its most simple form, is a method of combining data so as to extremize some quantity of interest. The optimality criterion depends on the particular application, but for signal processing, one typically wants to: (i) maximize the detection probability for a fixed rate of false alarms, (ii) maximize the signal-to-noise ratio of some test statistic, (iii) maximize the posterior distribution of some signal parameter, or (iv) find the minimal variance, unbiased estimator of some quantity. Finding such optimal combinations plays a key role in both Bayesian and frequentist approaches to statistical inference (Section 3), and it is an important tool for every data analyst. For a Bayesian, the optimal combinations are often implicitly contained in the likelihood function, while for a frequentist, optimal filtering is usually more explicit, as there is much more freedom in the construction of a statistic.

In this section, we give several simple examples of optimal (or matched) filtering for deterministic signals, and we then show how the standard optimally-filtered cross-correlation statistic [22, 26] for an Gaussian-stationary, unpolarized, isotropic gravitational-wave background can be derived as a matched-filter statistic for the expected cross-correlation. This derivation of the optimally-filtered cross-correlation statistic differs from the 'standard' one given e.g., in [22], but it illustrates a connection between searches for deterministic and stochastic signals, which is one of the goals of this review article.

6.1 Optimal combination of independent measurements

As a simple explicit example, suppose we have N *independent* measurements

$$d_i = a + n_i\,, \qquad i = 1, 2, \cdots, N\,, \tag{6.1}$$

where a is some astrophysical parameter that we want to estimate and n_i are (independent) noise terms. Assuming the noise has zero mean and known variance σ_i^2 (which can be different from measurement to measurement), it follows that

$$\langle d_i \rangle = a\,, \qquad \mathrm{Var}(d_i) \equiv \langle d_i^2 \rangle - \langle d_i \rangle^2 = \sigma_i^2\,. \tag{6.2}$$

The goal is to find a linear combination of the data

$$\hat{a} \equiv \sum_i \lambda_i d_i \tag{6.3}$$

that is optimal in the sense of being an *unbiased, minimal variance* estimator of a. Unbiased (i.e., $\langle \hat{a} \rangle = a$) implies

$$\sum_i \lambda_i = 1\,, \tag{6.4}$$

while minimum variance implies

$$\mathrm{Var}(\hat{a}) \equiv \sigma_{\hat{a}}^2 = \sum_i \lambda_i^2 \sigma_i^2 = \mathrm{minimum}\,. \tag{6.5}$$

Since (6.4) is a constraint that must hold when we minimize the variance, we can use Lagrange's method of undetermined multipliers [40] and minimize instead

$$f(\lambda_i, \Lambda) \equiv \sum_i \lambda_i^2 \sigma_i^2 + \Lambda \left(1 - \sum_i \lambda_i\right) \tag{6.6}$$

with respect to both λ_i and Λ. The final result is:

$$\lambda_i = \left(\sum_j \frac{1}{\sigma_j^2}\right)^{-1} \frac{1}{\sigma_i^2} \tag{6.7}$$

so that

$$\hat{a} = \left(\sum_j \frac{1}{\sigma_j^2}\right)^{-1} \sum_i \frac{d_i}{\sigma_i^2}\,. \tag{6.8}$$

Thus, the linear combination is a *weighted average* that gives less weight to the noiser measurements (i.e., those with large variance σ_i^2). The variance of the optimal combination is

$$\sigma_{\hat{a}}^2 = \left(\sum_j \frac{1}{\sigma_j^2}\right)^{-1}\,. \tag{6.9}$$

If the individual variances happen to be equal (i.e., $\sigma_i^2 \equiv \sigma^2$), then the above expressions reduce to $\hat{a} = N^{-1}\sum_i d_i$ and $\sigma_{\hat{a}}^2 = \sigma^2/N$, which are the standard formulas for the sample mean and the reduction in the variance for N independent and identically-distributed measurements as we saw in Section 3.5.

The above results can also be derived by maximizing the likelihood function

$$p(d|a, \sigma_1^2, \sigma_2^2, \cdots, \sigma_N^2) = \frac{1}{(2\pi)^{N/2}\sqrt{\sigma_1^2\sigma_2^2\cdots\sigma_N^2}} \exp\left[-\frac{1}{2}\sum_{i=1}^N \frac{(d_i - a)^2}{\sigma_i^2}\right] \tag{6.10}$$

with respect to the signal parameter a, assuming that the noise terms n_i are Gaussian-distributed and independent of one another. In fact, similar to what we showed in Section 3.5, one can rewrite the argument of the exponential so that

$$p(d|a, \sigma_1^2, \sigma_2^2, \cdots, \sigma_N^2) \propto \exp\left[-\frac{1}{2}\frac{(a - \hat{a})^2}{\sigma_{\hat{a}}^2}\right]\,, \tag{6.11}$$

where $\hat{a}$ and $\sigma_{\hat{a}}^2$ are given by (6.8) and (6.9), respectively. From this expression, it immediately follows that $\hat{a}$ maximizes the likelihood, and also the posterior distribution of a, if the prior for a is flat.

6.2 Correlated measurements

Suppose the N measurements d_i are *correlated*, so that the covariance matrix C has non-zero elements

$$C_{ij} \equiv \langle d_i d_j \rangle - \langle d_i \rangle \langle d_j \rangle \tag{6.12}$$

when $i \neq j$. Again, we want to find a linear combination (6.3) that is unbiased and has minimum variance

$$\sigma^2_{\hat{a}} = \sum_i \sum_j \lambda_i \lambda_j C_{ij} \,. \tag{6.13}$$

By following the same Lagrange multiplier procedure described in the previous subsection, one can show that the optimal estimator is

$$\hat{a} = \left(\sum_k \sum_l ||C^{-1}||_{kl} \right)^{-1} \sum_i \sum_j ||C^{-1}||_{ij} d_i \,. \tag{6.14}$$

Thus, the weighting factors $1/\sigma_i^2$ of the previous subsection are replaced by $\sum_j ||C^{-1}||_{ij}$. Note that for uncorrelated measurements, $C_{ij} = \delta_{ij}\sigma_i^2$, so the above expression for $\hat{a}$ reduces to that found previously in (6.8).

NOTE: Although (6.14) shows how to optimally combine data that are correlated with one another, it turns out that for most practical purposes one can get by using expressions like (6.8) and (6.18) below, which are valid for *uncorrelated* data. This is because the values of the Fourier transform of a stationary random process are uncorrelated for different frequency bins. Basically, the Fourier transform is a rotation in data space to a basis in which the covariance matrix is diagonal (this is called a *Karhunen-Loeve transformation*). This is one of the reasons why much of signal processing is done in the frequency domain.

6.3 Matched filter

Suppose that the astrophysical signal is not constant but also has a 'shape' h_i so that

$$d_i = a h_i + n_i \,, \qquad i = 1, 2, \cdots, N \,. \tag{6.15}$$

We will assume that the h_i are known, so that the only unknown signal parameter is a. We will also assume that the different measurements are independent, as will be the case for a stationary random process in the frequency domain. Since $\langle d_i \rangle = a h_i$ is not a constant, the analysis of the previous subsection does not immediately apply. However, if we simply rescale d_i by h_i, we obtain a new set of measurements

$$\bar{d}_i \equiv d_i / h_i \tag{6.16}$$

for which

$$\langle \bar{d}_i \rangle = a \,, \qquad \mathrm{Var}(\bar{d}_i) \equiv \bar{\sigma}_i^2 = \sigma_i^2 / h_i^2 \,, \tag{6.17}$$

so that the previous analysis *is* now valid. Thus,

$$\hat{a} = \left(\sum_j \frac{1}{\bar{\sigma}_j^2} \right)^{-1} \sum_i \frac{\bar{d}_i}{\bar{\sigma}_i^2} = \left(\sum_j \frac{h_j^2}{\sigma_j^2} \right)^{-1} \sum_i \frac{h_i d_i}{\sigma_i^2} \tag{6.18}$$

is the optimal estimator of a.

The above expression for $\hat{a}$ is often called a *matched filter* [195] since the data d_i are projected onto the expected signal shape h_i (as well as weighted by the inverse of the noise variance σ_i^2). The particular combination

$$Q_i \equiv h_i/\sigma_i^2 \tag{6.19}$$

multiplying d_i is the *optimal filter* for this analysis.[14] When there are many possible candidate signal shapes, one constructs a *template bank*—i.e., a collection of possible shapes against which the data compared. By normalizing each of the templates so that $\sum_i (h_i^2/\sigma_i^2) = 1$, the signal-to-noise ratio of the matched filter

$$\hat{\rho}(h) \equiv \sum_i \frac{h_i d_i}{\sigma_i^2} \tag{6.20}$$

can be used as a detection statistic. That is, the maximum value of $\hat{\rho}(h)$ over the space of templates $\{h_i\}$ is compared against some threshold ρ_* (chosen so that the false alarm probability is below some acceptable value). If the maximum signal-to-noise ratio exceeds the threshold, then one claims detection of the signal with a certain level of confidence. The shape of the detected signal is that which corresponds to the maximum signal-to-noise ratio.

6.4 Optimal filtering for a stochastic background

As noted by Fricke [68], the above results can be used to derive the optimal cross-correlation statistic for the stochastic background search. (A more 'standard' derivation can be found e.g., in [22].) To see this, consider a cross-correlation search for a Gaussian-stationary, unpolarized, isotropic gravitational-wave background using two detectors having uncorrelated noise. Let T be the total observation time of the measurement. In the frequency domain, the measurements are given by the values of the complex-valued cross-correlation

$$x(f) = \tilde{d}_1(f)\tilde{d}_2^*(f) \tag{6.21}$$

where $\tilde{d}_I(f)$, $I = 1, 2$ are the Fourier transforms of the time-series output of the two detectors:

$$\begin{aligned} d_1(t) &= h_1(t) + n_1(t)\,, \\ d_2(t) &= h_2(t) + n_2(t)\,. \end{aligned} \tag{6.22}$$

The $x(f)$ for different frequencies correspond to the measurements d_i of the previous subsections. Since we are assuming uncorrelated detector noise,

$$\langle x(f)\rangle = \langle \tilde{h}_1(f)\tilde{h}_2^*(f)\rangle = \frac{T}{2}\Gamma_{12}(f)S_h(f)\,, \tag{6.23}$$

[14] For correlated measurements, $Q_i = \sum_j ||\bar{C}^{-1}||_{ij}/h_i$ where $\bar{C}^{-1}$ is the inverse of the re-scaled covariance matrix $\bar{C}_{ij} \equiv C_{ij}/(h_i h_j)$.

where $S_h(f)$ is power spectral density of the stochastic background signal, and $\Gamma_{12}(f)$ is overlap reduction function for the two detectors.[15] In the weak-signal limit, the covariance matrix is dominated by the diagonal terms:

$$\begin{aligned} C_{ff'} &\equiv \langle x(f)x^*(f')\rangle - \langle x(f)\rangle\langle x^*(f')\rangle \\ &\approx \langle \tilde{n}_1(f)\tilde{n}_1^*(f')\rangle\langle \tilde{n}_2^*(f)\tilde{n}_2(f')\rangle \\ &= \frac{T}{4}P_{n_1}(f)P_{n_2}(f)\,\delta(f-f') \end{aligned} \tag{6.24}$$

where $P_{n_I}(f)$ are the 1-sided power spectral densities of the noise in the two detectors:

$$\langle \tilde{n}_I(f)\tilde{n}_I^*(f')\rangle = \frac{1}{2}P_{n_I}(f)\,\delta(f-f')\,. \tag{6.25}$$

Thus, in this approximation

$$\int_{-\infty}^{\infty} df'\,(C^{-1})_{ff'} \approx \frac{4}{T}\frac{1}{P_{n_1}(f)P_{n_2}(f)}\,. \tag{6.26}$$

Now, suppose we are searching for a stochastic background with a flat spectrum

$$\Omega_{\rm gw}(f) = \Omega_0 = {\rm const}\,, \tag{6.27}$$

whose amplitude Ω_0 we would like to estimate. Then, according to (2.17),

$$S_h(f) = \frac{3H_0^2}{2\pi^2}\frac{\Omega_0}{f^3} \equiv \Omega_0\bar{H}(f)\,, \tag{6.28}$$

so that

$$\frac{T}{2}\Gamma_{12}(f)\bar{H}(f) \quad \longleftrightarrow \quad h_i \tag{6.29}$$

is the expected signal 'shape' h_i in the notation of the previous subsections. Given (6.26) and (6.29), it is now a simple matter to show that

$$\hat{\Omega}_0 = \mathcal{N}\int_{-\infty}^{\infty} df\,\frac{\Gamma_{12}(f)\bar{H}(f)}{P_{n_1}(f)P_{n_2}(f)}\tilde{d}_1(f)\tilde{d}_2^*(f)\,, \tag{6.30}$$

where

$$\mathcal{N} \equiv \left[\frac{T}{2}\int_{-\infty}^{\infty} df\,\frac{\Gamma_{12}^2(f)\bar{H}^2(f)}{P_{n_1}(f)P_{n_2}(f)}\right]^{-1}\,. \tag{6.31}$$

The variance and expected signal-to-noise ratio of the estimator $\hat{\Omega}_0$ are:

$$\sigma^2_{\hat{\Omega}_0} = \left[T\int_{-\infty}^{\infty} df\,\frac{\Gamma_{12}^2(f)\bar{H}^2(f)}{P_{n_1}(f)P_{n_2}(f)}\right]^{-1}\,, \tag{6.32}$$

[15]The last equality in (6.23) follows from (5.41) with the Dirac delta function $\delta(f-f')$ replaced by its finite-time version $\delta_T(f-f') = T{\rm sinc}[\pi(f-f')T]$, which equals T when $f=f'$.

and

$$\rho^2 = \sqrt{T}\left[\int_{-\infty}^{\infty} df\ \frac{\Gamma_{12}^2(f) S_h^2(f)}{P_{n_1}(f) P_{n_2}(f)}\right]^{1/2}\ . \tag{6.33}$$

The combination

$$\tilde{Q}(f) \equiv \mathcal{N} \frac{\Gamma_{12}(f) \bar{H}(f)}{P_{n_1}(f) P_{n_2}(f)} \tag{6.34}$$

multiplying $\tilde{d}_1(f)\tilde{d}_2^*(f)$ in (6.30) is just the standard optimal filter derived in [22, 26]. The optimally-filtered cross-correlation statistic, denoted Y in those references, is given by $Y = \hat{\Omega}_0 T$.

7 Anisotropic backgrounds

Sameness is the mother of disgust, variety the cure. *Francesco Petrarch*

An anisotropic background of gravitational radiation has *preferred* directions on the sky—the associated signal is stronger coming from certain directions ("hot" spots) than from others ("cold" spots). The anisotropy is produced primarily by sources that follow the local distribution of matter in the universe (e.g., compact white-dwarf binaries in our galaxy), as opposed to sources at *cosmological* distances (e.g., cosmic strings or quantum fluctuations in the gravitational field amplified by inflation [22, 115]), which would produce an *isotropic* background. This means that the measured distribution of gravitational-wave power on the sky can be used to discriminate between cosmologically-generated backgrounds, produced in the very early Universe, and astrophysically-generated backgrounds, produced by more recent populations of astrophysical sources. In addition, an anisotropic distribution of power may allow us to detect the gravitational-wave signal in the first place; as the lobes of the antenna pattern of a detector sweep across the "hot" and "cold" spots of the anisotropic distribution, the amplitude of the signal is modulated in time, while the detector noise remains unaffected [16].

In this section, we describe several different approaches for searching for anisotropic backgrounds of gravitational waves: The first approach (described in Section 7.2) looks for modulations in the correlated output of a pair of detectors, at harmonics of the rotational or orbital frequency of the detectors (e.g., daily rotational motion for ground-based detectors like LIGO, Virgo, etc., or yearly orbital motion for space-based detectors like LISA). This approach assumes a known distribution of gravitational-wave power $\mathcal{P}(\hat{n})$, and filters the data so as to maximize the signal-to-noise ratio of the harmonics of the correlated signal. The second approach (Section 7.3) constructs maximum-likehood estimates of the gravitational-wave power on the sky based on cross-correlated data from a network of detectors. This approach produces sky maps of $\mathcal{P}(\hat{n})$, analogous to sky maps of temperature anisotropy in the cosmic microwave background radiation. The third approach (Section 7.4) constructs frequentist detection statistics for either an unknown or an assumed distribution of gravitational-wave power on the sky. The fourth and final approach we describe (Section 7.5) attempts to measure both the amplitude *and* phase of the gravitational-wave background at each point on the sky, making minimal assumptions about the statistical properties of the signal. This latter approach produces sky maps of the real and imaginary parts of the random fields $h_+(f,\hat{n})$ and $h_\times(f,\hat{n})$, from which the power in the background $\mathcal{P}(\hat{n}) = |h_+|^2 + |h_\times|^2$ is just one of many quantities that can be estimated from the measured data.

Numerous papers have been written over the last ≈ 20 years on the problem of detecting anisotropic stochastic backgrounds, starting with the seminal paper by Allen and Ottewill [25], which laid the foundation for much of the work that followed. Readers interested in more details should see [25] regarding modulations of the cross-correlation statistic at harmonics of the Earth's rotational frequency; [35, 34, 122, 175, 120, 166] for maximum-likelihood estimates of gravitational-wave power; [175, 164] for maximum-likelihood ratio detection statistics; and [69, 51, 142] regarding phase-coherent mapping. For results of actual analyses of initial LIGO data and pulsar timing data for anisotropic backgrounds, see [5, 169] and Section 10.2.5.

Note that we will not discuss in any detail methods to detect anisotropic backgrounds using space-based interferometers like LISA or the Big-Bang Observer (BBO). As mentioned in Section 5.5.2, the confusion noise from the galactic population of compact white dwarf binaries is a guaranteed source of anisotropy for such detectors. At low frequencies, measurements made using a single LISA will be sensitive to only the $l = 0, 2, 4$ components of the background, while cross-correlating data from two independent LISA-type detectors (as in BBO) will allow for extraction of the full range of multipole moments. The proposed data analysis methods are similar to those that we will discuss in Sections 7.2 and 7.3, but using the synthesized A, E, and T data channels for a single LISA (see Section 9.6). Readers should see [74, 45, 183, 151, 154, 105, 60, 165] for details.

7.1 Preliminaries

7.1.1 Quadratic expectation values

For simplicity, we will restrict our attention to Gaussian-stationary, unpolarized, anisotropic backgrounds with quadratic expectation values given by (2.15):

$$\langle h_A(f,\hat{n}) h^*_{A'}(f',\hat{n}')\rangle = \frac{1}{4}\mathcal{P}(f,\hat{n})\delta(f-f')\delta_{AA'}\delta^2(\hat{n},\hat{n}')\,, \tag{7.1}$$

where

$$S_h(f) = \int d^2\Omega_{\hat{n}}\, \mathcal{P}(f,\hat{n})\,. \tag{7.2}$$

We will also assume that $\mathcal{P}(f,\hat{n})$ factorizes

$$\mathcal{P}(f,\hat{n}) = \bar{H}(f)\mathcal{P}(\hat{n})\,, \tag{7.3}$$

so that the angular distribution of power on the sky is independent of frequency. We will chose our normalization so that $\bar{H}(f_R) = 1$, where f_R is a reference frequency, typically taken to equal 100 Hz for ground-based detectors. We will also assume that the spectral shape $\bar{H}(f)$ is known, so that we only need to recover $\mathcal{P}(\hat{n})$. If we expand the power $\mathcal{P}(\hat{n})$ in terms of spherical harmonics,

$$\mathcal{P}(\hat{n}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \mathcal{P}_{lm} Y_{lm}(\hat{n})\,, \tag{7.4}$$

then this normalization choice is equivalent to $\mathcal{P}_{00} = S_h(f_R)/\sqrt{4\pi}$, and has units of $(\text{strain})^2\,\text{Hz}^{-1}\,\text{rad}^{-2}$. Thus, $\mathcal{P}_{00}$ is a measure of the *isotropic* component of the background, and sets the overall normalization of the strain power spectral density $S_h(f)$.

7.1.2 Short-term Fourier transforms

Since the response of a detector will changes as its antenna pattern sweeps across the "hot" and "cold" spots of an anisotropic distribution, we will need to split the data taken by the detectors into chunks of duration τ, where τ is much greater than the light-travel time between any pair of detectors, but small enough that the detector response functions do not change appreciably over that interval. (For Earth-based interferometers like LIGO, $\tau \sim 100$ s to 1000 s is appropriate.) Each chunk of data $[t - \tau/2, t + \tau/2]$ will then be Fourier transformed over the duration τ, yielding

$$\tilde{d}_I(t; f) = \int_{t-\tau/2}^{t+\tau/2} dt' \, d_I(t') e^{-i2\pi f t'} \, . \tag{7.5}$$

This operation is often called a *short-term* Fourier transform. Note that, in this notation, t labels a *particular* time chunk, and is not a variable that is subsequently Fourier transformed.

7.1.3 Cross-correlations

For many of the approaches that map the distribution of gravitational-wave power, it is convenient to work with cross-correlated data from two detectors, evaluated at the same time chunk t and frequency f:

$$\hat{C}_{IJ}(t; f) = \frac{2}{\tau} \tilde{d}_I(t; f) \tilde{d}_J^*(t; f) \, . \tag{7.6}$$

The factor of 2 is a convention consistent with the choice of one-sided power spectra. Assuming uncorrelated detector noise and using expectation values given in (7.1), we find

$$\langle \hat{C}_{IJ}(t; f) \rangle = \bar{H}(f) \int d^2\Omega_{\hat{n}} \, \gamma_{IJ}(t; f, \hat{n}) \mathcal{P}(\hat{n}) \, , \tag{7.7}$$

where

$$\gamma_{IJ}(t; f, \hat{n}) \equiv \frac{1}{2} \sum_A R_I^A(t; f, \hat{n}) R_J^{A*}(t; f, \hat{n}) \, . \tag{7.8}$$

Note that up to a factor of $1/(4\pi)$, the function $\gamma_{IJ}(t; , f, \hat{n})$ is just the integrand of the isotropic overlap function $\Gamma_{IJ}(f)$ given by (5.39). In what follows, we will drop the detector labels IJ from both $\hat{C}_{IJ}(t; f)$ and $\gamma_{IJ}(t; f, \hat{n})$ when there is no chance for confusion.

Figure 42 shows maps of the real and imaginary parts of $\gamma(t; f, \hat{n})$ (appropriately normalized) for the strain response of the 4-km LIGO Hanford and LIGO Livingston interferometers evaluated at $f = 0$ Hz (top two plots) and $f = 200$ Hz (bottom two plots). (In the Earth-fixed frame, the detectors don't move so there is no time dependence to worry about.) Note the presence of oscillations or 'lobes' for the $f = 200$ Hz plots, which come from the exponential factor $e^{-i2\pi f \hat{n} \cdot \Delta\vec{x}/c}$ of the product of the two response functions (5.46). For $f = 0$, this factor is unity.

Figure 43 is a similar plot, showing Mollweide projections of $\gamma(t; f, \hat{n})$ for the Earth-term-only Doppler frequency response (5.22) of pairs of pulsars separated on the sky by $\zeta = 0°$, $45°$, $90°$, $135°$, $180°$. (There is no time dependence nor frequency dependence for these functions.) The bottom panel is a plot of the Hellings and Downs curve as a function of the angular separation between a pair of Earth-pulsar baselines. By integrating the top plots over the whole sky (appropriately normalized), one obtains the values of the Hellings and Downs curve for those angular separations.

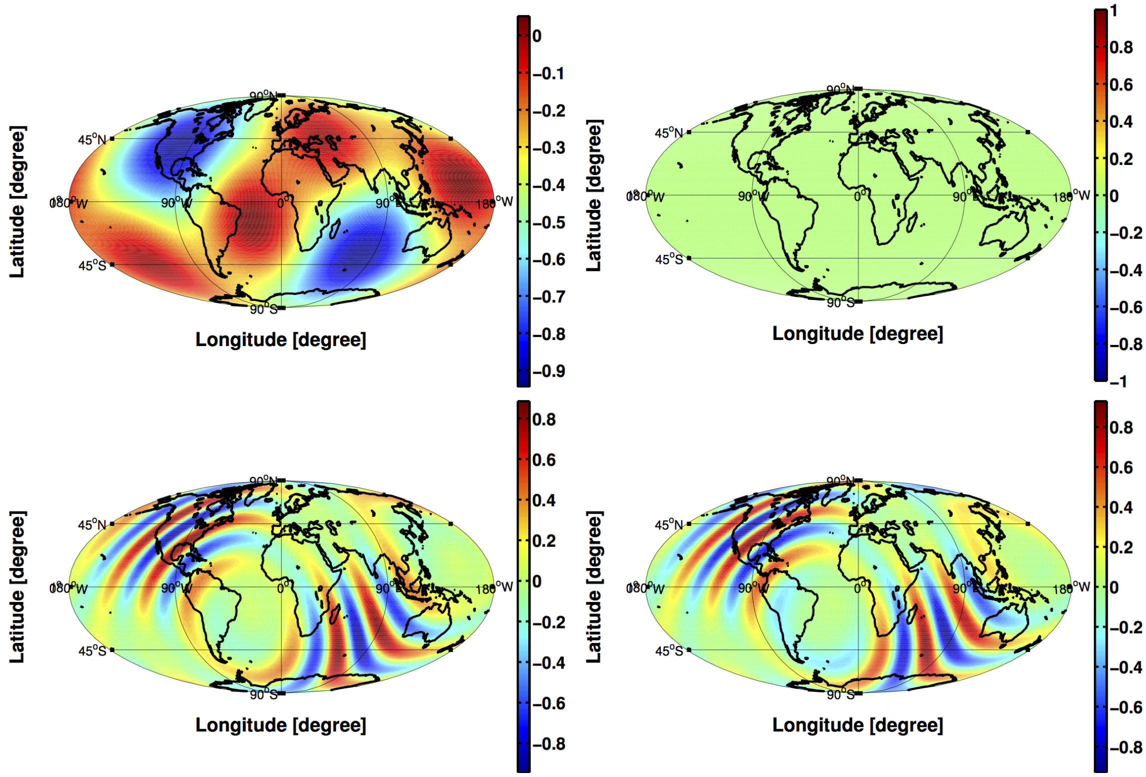


Figure 42: Real and imaginary parts of $\gamma(f, \hat{n})$ (appropriately normalized) for the strain response of the 4-km LIGO Hanford and LIGO Livingston interferometers for $f = 0$ Hz (top two plots) and $f = 200$ Hz (bottom two plots). In the top left plot, note the large blue region in the vicinity of the two detectors, corresponding to the *anti-alignment* of the Hanford and Livingston interferometers—i.e., the arms of the two interferometers are rotated by 90° with respect to one another. As shown in the top right plot, there is no imaginary component to the integrand of the overlap function at 0 Hz. The bottom two plots show multiple positive and negative oscillations ('lobes'), which come from the exponential factor $e^{-i2\pi f\hat{n}\cdot\Delta\vec{x}/c}$ of the product of the two response functions (5.46). The location of the positive and negative lobes are shifted relative to one another for the real and imaginary parts. The separation between the lobes depends inversely on the frequency.

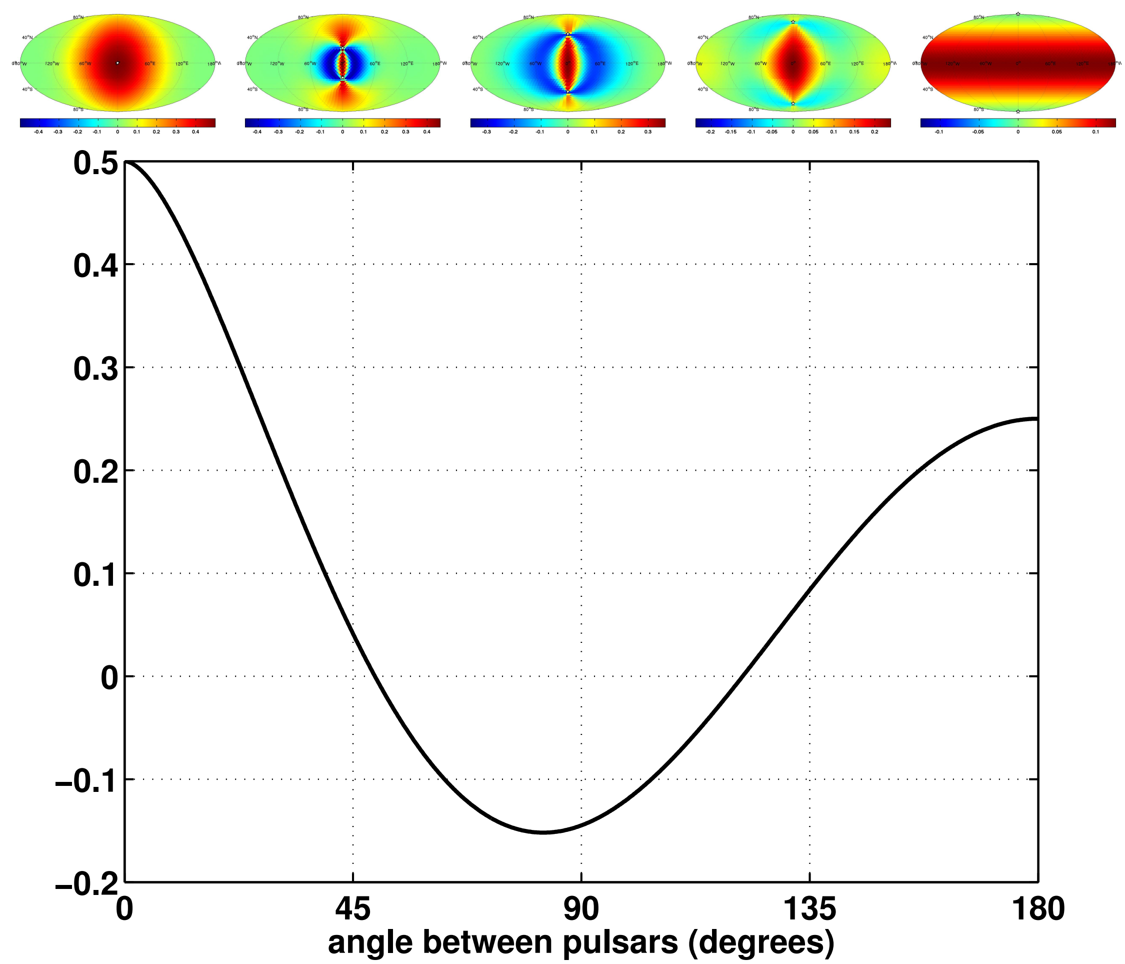


Figure 43: Top row: Mollweide projections of $\gamma(\hat{n})$ for pairs of pulsars separated on the sky by $\zeta = 0°$, $45°$, $90°$, $135°$, $180°$. Reddish regions correspond to positive values of $\gamma(\hat{n})$; blueish regions correspond to negative values of $\gamma(\hat{n})$. Bottom: Hellings and Downs curve as a function of the angular separation between two distinct pulsars. The integral of the top plots over the whole sky equal the values of the Hellings and Downs curve for these angular separations. (See also Figure 36.)

7.1.4 Spherical harmonic components of $\gamma(t; f, \hat{n})$

As first noted in [25], the functions $\gamma(t; f, \hat{n})$ defined above (7.8) play a very important role in searches for anisotropic backgrounds. For a fixed pair of detectors at a fixed time t and for fixed frequency f, these functions are scalar fields on the unit 2-sphere and hence can be expanded in terms of the ordinary spherical harmonics $Y_{lm}(\hat{n})$:

$$\gamma(t; f, \hat{n}) \equiv \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \gamma_{lm}(t; f) Y_{lm}^*(\hat{n}) \,, \tag{7.9}$$

or, equivalently,

$$\gamma_{lm}(t; f) \equiv \int d^2\Omega_{\hat{n}} \, \gamma(t; f, \hat{n}) Y_{lm}(\hat{n}) \,. \tag{7.10}$$

Note that this definition differs from (7.4) for $\mathcal{P}_{lm}$ by a complex conjugation, but agrees with the convention used in [25]. In terms of the spherical harmonic components, it follows that

$$\int d^2\Omega_{\hat{n}} \, \gamma(t; f, \hat{n}) \mathcal{P}(\hat{n}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \gamma_{lm}(t; f) \mathcal{P}_{lm} \,, \tag{7.11}$$

as a consequence of the orthogonality of the $Y_{lm}(\hat{n})$. This expression enters (7.7) for the expected cross-correlation of the output in two detectors. As explained in [25, 175], the time dependence of $\gamma_{lm}(t; f)$ is particularly simple:

$$\gamma_{lm}(t; f) = \gamma_{lm}(0; f) \, e^{im 2\pi t / T_{\rm mod}} \,, \tag{7.12}$$

where $T_{\rm mod}$ is the relevant modulation period associated with the motion of the detectors. For example, for ground-based detectors like LIGO and Virgo, $T_{\rm mod} = 1$ sidereal day, since the displacement vector $\Delta\vec{x}(t) \equiv \vec{x}_2(t) - \vec{x}_1(t)$ connecting the vertices of the two interferometers (and which enters the expression for the overlap function) traces out a cone on the sky with a period of one sidereal day. If there is no time dependence, as is the case for pulsar timing, $T_{\rm mod}$ is infinite.

Example: Earth-based interferometers

As was also shown in [25], one can derive *analytic* expressions for $\gamma_{lm}(t; f)$ for a pair of Earth-based interferometers in the short-antenna limit. If we set $t = 0$, then $\gamma_{lm}(0; f)$ can be written as a linear combination[16] involving spherical Bessel functions, $j_n(x)/x^n$ (for l even) and $j_n(x)/x^{n-1}$ (for l odd), where x depends on the relative separation of the two detectors, $x \equiv 2\pi f |\Delta\vec{x}|/c$. The coefficients of the expansions are complex numbers that depend on the relative orientation of the detectors. Explicit expression for the first few spherical harmonic components for the LIGO Hanford-LIGO Livingston pair are given

[16] The number of terms in the expansion is given by $2 + \text{floor}(1 + L/2)$.

below:

$$
\begin{aligned}
\gamma_{00}(0;f) &= -\,0.0766 j_0(x) - 2.1528 j_1(x)/x + 2.4407 j_2(x)/x^2\,,\\
\gamma_{10}(0;f) &= -\,0.0608i\, j_1(x) - 2.6982i\, j_2(x)/x + 7.7217i\, j_3(x)/x^2\,,\\
\gamma_{11}(0;f) &= -\,(0.0519 + 0.0652i) j_1(x) - (1.8621 + 1.0517i) j_2(x)/x\\
&\quad + (4.0108 - 2.4933i) j_3(x)/x^2\,,\\
\gamma_{20}(0;f) &= 0.0316 j_0(x) - 0.9612 j_1(x)/x + 10.9038 j_2(x)/x^2 - 52.7905 j_3(x)/x^3\,,\\
\gamma_{21}(0;f) &= -\,(0.0669 - 0.0532i) j_0(x) - (1.9647 - 2.6145i) j_1(x)\\
&\quad + (15.0524 - 24.7604i) j_2(x)/x^2 - (36.5620 - 50.7179i) j_3/x^3\,,\\
\gamma_{22}(0;f) &= -\,(0.0186 - 0.0807i) j_0(x) + (1.2473 + 1.6858i) j_1(x)/x\\
&\quad - (12.2048 + 12.5814i) j_2(x)/x^2 + (60.7859 + 12.7191i) j_3(x)/x^3\,.
\end{aligned}
\tag{7.13}
$$

Note that the above numerical coefficients do not agree with those in [25] due to an overall normalization factor of $4\pi/5$ and phase $e^{im\phi}$, where $\phi = -38.52°$ is the angle between the separation vector between the vertices of the LIGO-Hanford and LIGO-Livingston interferometers and the Greenwich meridian [175]. Plots of the real and imaginary parts of $\gamma_{lm}(0;f)$ for $l = 0,\ 1,\ 2,\ 3,\ 4$ and $m \geq 0$ for the the LIGO Hanford-LIGO Livingston detector pair are given in Figure 44. For $m < 0$, one can use the relation

$$
\gamma_{lm}(t;f) = (-1)^{l+m}\gamma_{l,-m}(t;f)\,, \tag{7.14}
$$

which follows from the properties of the spherical harmonics $Y_{lm}(\hat{n})$ (see Appendix D). Note that up to an overall normalization factor of $5/\sqrt{4\pi}$, the real part of $\gamma_{00}(0;f)$ is the Hanford-Livingston overlap function for an unpolarized, isotropic stochastic background, shown in Figure 33.

Example: Pulsar timing arrays

In Figure 45, we show plots of the spherical harmonic components of $\gamma(t;f,\hat{n})$ calculated using the Earth-term-only Doppler-frequency response functions (5.22) for pulsar timing. Since there is no frequency or time-dependence for these response functions, the spherical harmonic components of $\gamma(\hat{n})$ depend only of the angular separation ζ between the two pulsars that define the detector pair. As shown in [120, 69], these functions can be calculated analytically for *all* values of l and m. A detailed derivation with all the relevant formulae can be found in Appendix E of [69]; there the calculation is done in a 'computational' frame, where one of the pulsars is located along the z-axis and the other is in the xz-plane, making an angle ζ with respect to the first. In this computational frame, all of the components $\gamma_{lm}(\zeta)$ are real. Note that up to an overall normalization factor[17] of $3/\sqrt{4\pi}$, the function $\gamma_{00}(\zeta)$ is just the Hellings and Downs function for an unpolarized, isotropic stochastic background, shown in Figure 36.

[17] The functions here are a factor of 1/2 smaller than those in Figure 8 in [69], due to different definitions of $\gamma(t;f,\hat{n})$. Compare (115) in that paper to (7.8) and (7.10) above.

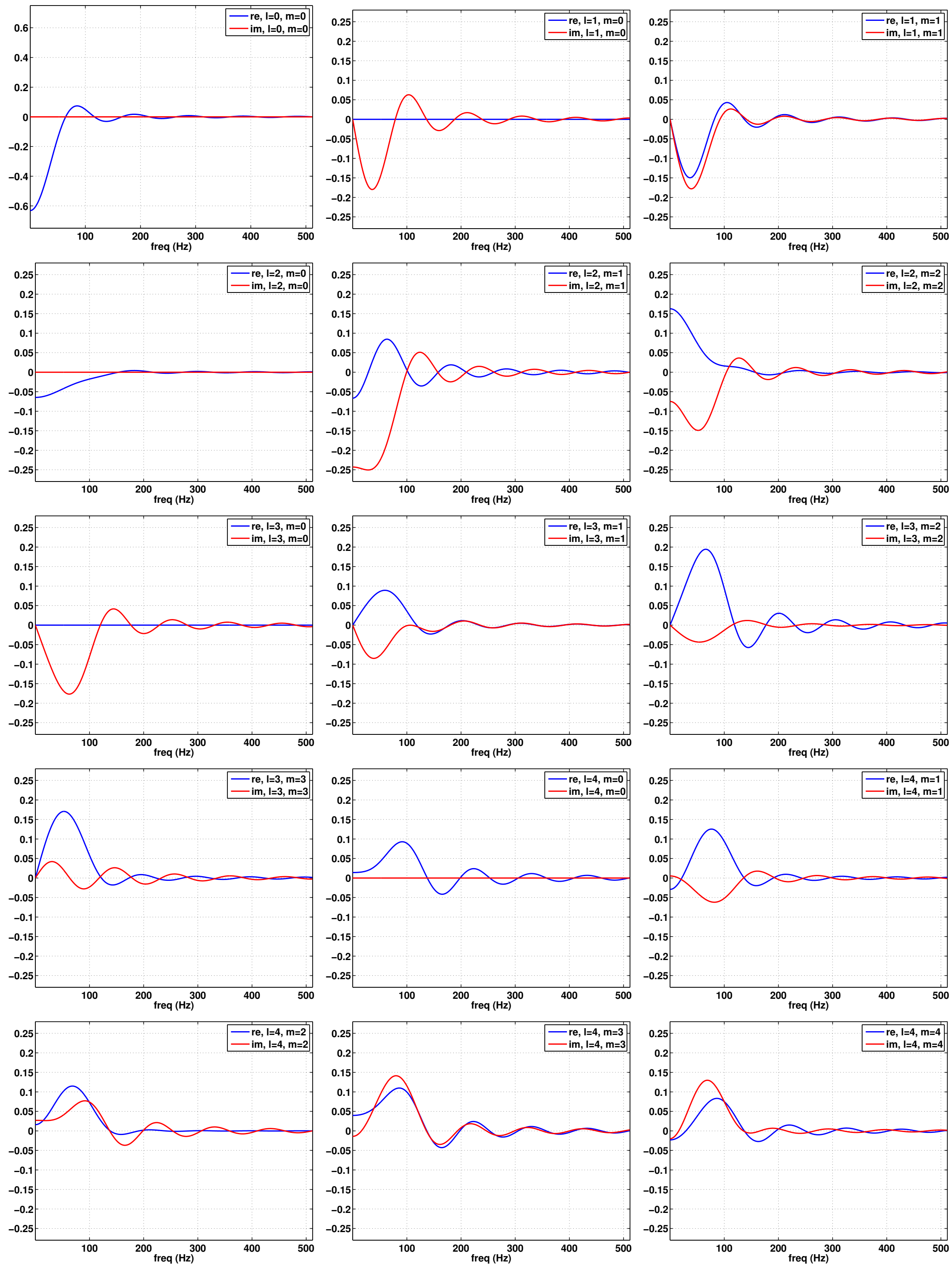


Figure 44: Real and imaginary parts of the spherical harmonic components $\gamma_{lm}(0; f)$ for the LIGO Hanford-LIGO Livingston detector pair. Here we show plots for $l = 0$, 1, 2, 3, 4 and $m \geq 0$. For $m < 0$, use (7.14).

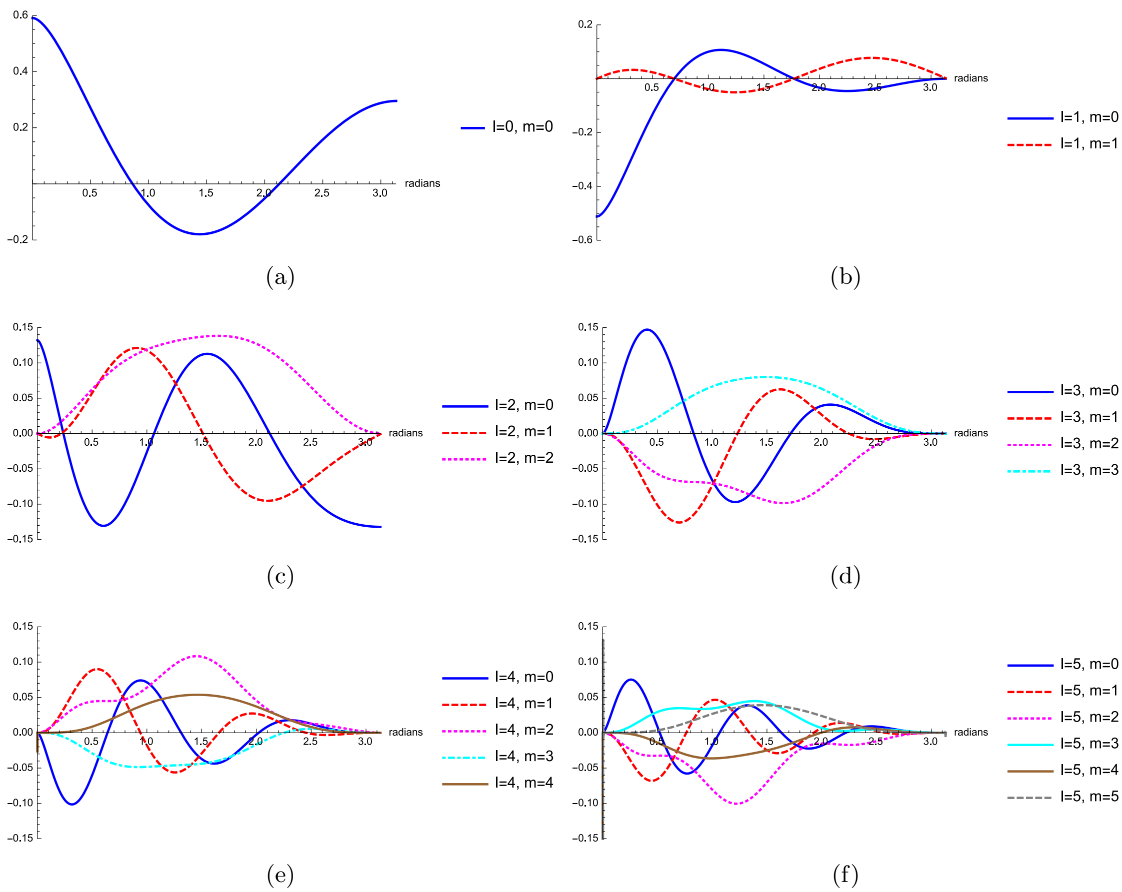


Figure 45: Spherical harmonic component functions $\gamma_{lm}(\zeta)$ for pulsar timing as a function of the angle ζ between two distinct pulsars. Here we show plots for $l = 0, 1, \cdots, 5$ and $m \geq 0$. We used the Earth-term-only Doppler-frequency response (5.22) to calculate these functions.

7.2 Modulations in the correlated output of two detectors

For ground-based detectors like LIGO and Virgo, an anisotropic gravitational-wave background will modulate the correlated output of a pair of detectors at harmonics of the Earth's rotational frequency. It turns out that for an unpolarized, anisotropic background, the contribution to the mth harmonic of the correlation has a frequency dependence proportional to

$$\bar{H}(f)\sum_{l=|m|}^{\infty}\gamma_{lm}(0;f)\mathcal{P}_{lm}\,, \tag{7.15}$$

where $\mathcal{P}_{lm}$ are the spherical harmonic components of the gravitational-wave power on the sky $\mathcal{P}(\hat{n})$. (We are assuming here that the spherical harmonic decomposition of $\mathcal{P}(\hat{n})$ is with respect to a coordinate system whose z-axis points along the Earth's rotational axis.) In this section, we derive the above result following the presentation in [25] and construct an optimal filter for the cross-correlation that maximizes the signal-to-noise ratio for the mth harmonic. This was the first concrete approach that was proposed for detecting an anisotropic stochastic background.

7.2.1 Time-dependent cross-correlation

We start by writing down an expression (in the frequency domain) for the correlated output of two ground-based detectors (e.g., LIGO Hanford and LIGO Livingston):

$$\hat{C}(t)=\int_{-\infty}^{\infty}df\,\tilde{Q}(f)\tilde{d}_1(t;f)\tilde{d}_2^*(t;f)\,, \tag{7.16}$$

where $\tilde{d}_{1,2}(t;f)$ are (short-term) Fourier transforms (7.5) centered around t, and where we have included a filter function $\tilde{Q}(f)$, whose specific form we will specify later. Since the cross-correlation is periodic with a period $T_{\rm mod}=1$ sidereal day (due to the motion of the detectors attached to the surface of the Earth), we can expand $\hat{C}(t)$ as a Fourier series:

$$\begin{aligned}\hat{C}(t)&=\sum_{m=-\infty}^{\infty}\hat{C}_m e^{im2\pi t/T_{\rm mod}}\,,\\ \hat{C}_m&=\frac{1}{T}\int_0^T dt\,\hat{C}(t)e^{-im2\pi t/T_{\rm mod}}\,.\end{aligned} \tag{7.17}$$

Here T is the total observation time, e.g., 1 sidereal year, which we will assume for simplicity is an integer multiple of $T_{\rm mod}$.

Assuming as usual that the detector noise is uncorrelated across detectors, and using the expectation values (7.1) for an unpolarized, anisotropic background, we find

$$\langle\hat{C}(t)\rangle=\frac{\tau}{2}\int_{-\infty}^{\infty}df\,\tilde{Q}(f)\bar{H}(f)\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\gamma_{lm}(t;f)\mathcal{P}_{lm}\,, \tag{7.18}$$

where $\gamma_{lm}(t;f)$ are the spherical harmonic components of $\gamma_{12}(t;f,\hat{n})$. (We have dropped the 12 indices to simplify the notation.) Similarly, if we assume that the gravitational-wave signal is weak compared to the detector noise, and that the duration τ is also much larger than the correlation time of the detectors, then

$$\langle \hat{C}(t)\hat{C}^*(t')\rangle - \langle \hat{C}(t)\rangle\langle \hat{C}^*(t')\rangle \approx \frac{\tau}{4}\delta^2_{tt'}\int_{-\infty}^{\infty} df\, |\tilde{Q}(f)|^2 P_{n_1}(f)P_{n_2}(f)\,, \tag{7.19}$$

where $P_{n_I}(f)$ is the one-sided power spectral density for the noise in detector $I = 1, 2$. These two results can now be cast in terms of the Fourier components $\hat{C}_m$ using (7.17). Since (7.12) implies

$$\frac{1}{T}\int_0^T dt\, \gamma_{lm'}(t; f)e^{-im2\pi t/T_{\rm mod}} = \delta_{mm'}\,\gamma_{lm}(0; f)\,, \tag{7.20}$$

we immediately obtain

$$\langle \hat{C}_m\rangle = \frac{\tau}{2}\int_{-\infty}^{\infty} df\, \tilde{Q}(f)\bar{H}(f)\sum_{l=|m|}^{\infty}\gamma_{lm}(0; f)\mathcal{P}_{lm}\,, \tag{7.21}$$

where we used

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l} = \sum_{m=-\infty}^{\infty}\sum_{l=|m|}^{\infty}\,. \tag{7.22}$$

Similarly,

$$\langle \hat{C}_m\hat{C}^*_{m'}\rangle - \langle \hat{C}_m\rangle\langle \hat{C}^*_{m'}\rangle \approx \delta_{mm'}\frac{1}{T}\left(\frac{\tau}{2}\right)^2\int_{-\infty}^{\infty} df\, |\tilde{Q}(f)|^2 P_{n_1}(f)P_{n_2}(f) \tag{7.23}$$

for the covariance of the estimators.

7.2.2 Calculation of the optimal filter

To determine the optimal form of the filter $\tilde{Q}(f)$ for the mth harmonic $\hat{C}_m$, we *maximize* the (squared) signal-to-noise:

$$\mathrm{SNR}^2_m \equiv \frac{|\langle C_m\rangle|^2}{\langle|\hat{C}_m|^2\rangle - |\langle\hat{C}_m\rangle|^2} = \frac{T\left|\int_{-\infty}^{\infty} df\, \tilde{Q}(f)\bar{H}(f)\sum_{l=|m|}^{\infty}\gamma_{lm}(0; f)\mathcal{P}_{lm}\right|^2}{\int_{-\infty}^{\infty} df\, |\tilde{Q}(f)|^2 P_{n_1}(f)P_{n_2}(f)}\,. \tag{7.24}$$

The above expression can be written in a more suggestive form if we introduce an *inner product* on the space of complex-valued functions [22]:

$$(A, B) \equiv \int_{-\infty}^{\infty} df\, A(f)B^*(f)P_{n_1}(f)P_{n_2}(f)\,. \tag{7.25}$$

In terms of this inner product,

$$\mathrm{SNR}^2_m = \frac{T\left|\left(\tilde{Q}, \frac{\bar{H}}{P_{n_1}P_{n_2}}\sum_{l=|m|}^{\infty}\gamma_{lm}\mathcal{P}_{lm}\right)\right|^2}{(\tilde{Q}, \tilde{Q})}\,. \tag{7.26}$$

But now the maximization problem is trivial, as it has been cast as a simple problem in vector algebra—namely to find the vector $\tilde{Q}$ that maximizes the ratio $|(\tilde{Q},A)|^2/(\tilde{Q},\tilde{Q})$ for a *fixed* vector A. But since this ratio is proportional to the squared cosine of the angle between $\tilde{Q}$ and A, it is maximized by choosing $\tilde{Q}$ *proportional* to A. Thus,

$$\tilde{Q}(f) \propto \frac{\bar{H}(f)}{P_{n_1}(f)P_{n_2}(f)} \sum_{l=|m|}^{\infty} \gamma_{lm}(0;f)\mathcal{P}_{lm} \tag{7.27}$$

is the form of the filter function that maximizes the SNR for the mth harmonic.

Note that this expression reduces to the standard form of the optimal filter (6.34) for an isotropic background, $\mathcal{P}_{lm} = \delta_{l0}\delta_{m0}\mathcal{P}_{00}$. Note also that the optimal filter assumes knowledge of both the spectral shape $\bar{H}(f)$ *and* the angular distribution of gravitational-wave power on the sky, $\mathcal{P}_{lm}$. So if one has some model for the expected anisotropy (e.g., a dipole in the same direction as the cosmic microwave background), then one can filter the cross-correlated data to be optimally sensitive to the harmonics $\hat{C}_m$ induced by the anisotropy.

7.2.3 Inverse problem

In [25], there was no attempt to solve the 'inverse problem'—that is, given the *measured values* of the correlation harmonics, how can one *infer* (or *estimate*) the components $\mathcal{P}_{lm}$? The first attempt to solve the inverse problem was given in [45], in the context of correlation measurements for both ground-based and space-based interferometers. Further developments in solving the inverse problem were given in subsequent papers, e.g., [34, 35, 122, 175], which we explain in more detail in the following subsections. Basically, these latter methods constructed frequentist maximum-likelihood estimators for the $\mathcal{P}_{lm}$, using singular-value decomposition to 'invert' the Fisher matrix (or point spread function), which maps the true gravitational-wave power distribution to the measured distribution on the sky.

7.3 Maximum-likelihood estimates of gravitational-wave power

In this section, we describe an approach for constructing maximum-likehood estimates of the gravitational-wave power distribution $\mathcal{P}(\hat{n})$. It is a solution to the inverse problem discussed at the end of the previous subsection. But since a network of gravitational-wave detectors typically does not have perfect coverage of the sky, the inversion requires some form of regularization, which we describe below. The gravitational-wave radiometer and spherical harmonic decomposition methods (Section 7.3.6) are the two main implementations of this approach, and have been used to analyze LIGO science data [5].

7.3.1 Likelihood function and maximum-likelihood estimators

As shown in Section 7.1.3 the cross-correlated data from two detectors

$$\hat{C}_{IJ}(t;f) = \frac{2}{\tau}\tilde{d}_I(t;f)\tilde{d}_J^*(t;f)\,. \tag{7.28}$$

has expectation values

$$\langle \hat{C}_{IJ}(t;f)\rangle = \bar{H}(f)\int d^2\Omega_{\hat{n}}\,\gamma_{IJ}(t;f,\hat{n})\mathcal{P}(\hat{n})\,. \tag{7.29}$$

We can write this relation abstractly as a matrix equation

$$\langle \hat{C}_{IJ}\rangle = M_{IJ}\,\mathcal{P}\,, \tag{7.30}$$

where $M_{IJ}\equiv \bar{H}(f)\gamma_{IJ}(t;f,\hat{n})$ and the matrix product is summation over directions $\hat{n}$ on the sky. The covariance matrix for the cross-correlated data is given by

$$\begin{aligned} N_{tf,t'f'} &\equiv \langle \hat{C}_{IJ}(t,f)\hat{C}^*_{IJ}(t',f')\rangle - \langle \hat{C}_{IJ}(t,f)\rangle\langle \hat{C}^*_{IJ}(t',f')\rangle \\ &\approx \delta_{tt'}\delta_{ff'}P_{n_I}(t;f)P_{n_J}(t;f)\,, \end{aligned} \tag{7.31}$$

where we have assumed as before that there is no cross-correlated detector noise, and that the gravitational-wave signal is weak compared to the detector noise.

If we treat the detector noise and the gravitational-wave spectral shape $\bar{H}(f)$ as known quantities (or if we estimate the detector noise from the auto-correlated output of each detector), then we can write down a likelihood function for the cross-correlated data given the signal model (7.30). Assuming a Gaussian-stationary distribution for the noise, we have

$$p(\hat{C}|\mathcal{P}) \propto \exp\left[-\frac{1}{2}(\hat{C}-M\mathcal{P})^\dagger N^{-1}(\hat{C}-M\mathcal{P})\right]\,, \tag{7.32}$$

where we have temporarily dropped the IJ indices for notational convenience.[18] Since the gravitational-wave power distribution $\mathcal{P}$ enters quadratically in the exponential of the likelihood, we can immediately write down the maximum-likelihood estimators of $\mathcal{P}$:

$$\hat{\mathcal{P}} = F^{-1}X\,, \tag{7.33}$$

where

$$F\equiv M^\dagger N^{-1}M\,,\qquad X\equiv M^\dagger N^{-1}\hat{C}\,. \tag{7.34}$$

The (square) matrix F is called the *Fisher information matrix*. It is typically a singular matrix, since the response matrix $M=\bar{H}\gamma$ usually has *null* directions (i.e., anisotropic distributions of gravitational-wave power that are mapped to zero by the detector response). Inverting F therefore requires some sort of regularization, such as singular value decomposition [136] (Section 7.3.5). The vector X is the so-called *dirty map*, as it represents the gravitational-wave sky as 'seen' by a pair of detectors. If the spectral shape $\bar{H}(f)$ that we used for our signal model exactly matches that of the observed background, then

$$\langle X\rangle = M^\dagger N^{-1}M\,\mathcal{P} = F\,\mathcal{P}\,. \tag{7.35}$$

Thus, even in the absence of noise, a point source $\mathcal{P}(k)=\delta^2(\hat{n},\hat{n}_0)$ does not map to a point source by the response of the detectors, but it maps instead to $F_{\hat{n}\hat{n}_0}$. This 'blurring' or 'spreading' of point sources is represented by a *point spread function*, which is a characteristic feature of any imaging system. We give plots of point spread functions for both pulsar timing arrays and ground-based interferometers in Section 7.3.4.

[18] The multiplications inside the exponential are *matrix* multiplications—either summations over sky directions $\hat{n}$ or summations over discrete times and frequencies, t and f.

7.3.2 Extension to a network of detectors

The above results generalize to a *network* of detectors. One simply replaces X and F in (7.33) by their network expressions, which are simply sums of the the dirty maps and Fisher matrices for each distinct detector pair:

$$X = \sum_I \sum_{J>I} X_{IJ}\,, \qquad F = \sum_I \sum_{J>I} F_{IJ}\,. \tag{7.36}$$

Explicit expressions for the dirty map and Fisher matrix for a network of detectors are:

$$X \equiv X_{\hat{n}} = \sum_I \sum_{J>I} \sum_t \sum_f \gamma^*_{IJ}(t; f, \hat{n}) \frac{\bar{H}(f)}{P_{n_I}(t; f) P_{n_J}(t, f)} \hat{C}_{IJ}(t; f)\,, \tag{7.37}$$

and

$$F \equiv F_{\hat{n}\hat{n}'} = \sum_I \sum_{J>I} \sum_t \sum_f \gamma^*_{IJ}(t; f, \hat{n}) \frac{\bar{H}^2(f)}{P_{n_I}(t; f) P_{n_J}(t, f)} \gamma_{IJ}(t; f, \hat{n}')\,. \tag{7.38}$$

Note that including more detectors in the network is itself a form of regularization, as adding more detectors typically means better coverage of the sky. This tends to 'soften' the singularities that may exist when trying to deconvolve (i.e., invert) the detector response.

7.3.3 Error estimates

Using (7.35) it follows that $\hat{\mathcal{P}}$ is an unbiased estimator of $\mathcal{P}$:

$$\langle \hat{\mathcal{P}} \rangle = \mathcal{P}\,. \tag{7.39}$$

Similarly, in the weak-signal approximation,

$$\begin{aligned} \langle X X^\dagger \rangle - \langle X \rangle \langle X^\dagger \rangle &\approx F\,, \\ \langle \hat{\mathcal{P}} \hat{\mathcal{P}}^\dagger \rangle - \langle \hat{\mathcal{P}} \rangle \langle \hat{\mathcal{P}}^\dagger \rangle &\approx F^{-1}\,. \end{aligned} \tag{7.40}$$

Thus, F is the covariance matrix for the dirty map X, while F^{-1} is the covariance matrix of the clean map $\hat{\mathcal{P}}$. We will see below (Section 7.3.5) that regularization necessarily changes these results as one cannot recover modes of $\mathcal{P}$ to which the detector network is insensitive. This introduces a bias in $\hat{\mathcal{P}}$, and changes the corresponding elements of the covariance matrix for $\hat{\mathcal{P}}$.

7.3.4 Point spread functions

As discussed in the previous section, the point spread function for mapping gravitational-wave power is given by the components of the Fisher information matrix:

$$\mathrm{PSF}_{\hat{n}_0}(\hat{n}) \equiv \mathrm{PSF}(\hat{n}, \hat{n}_0) = F_{\hat{n}\hat{n}_0}\,. \tag{7.41}$$

Here $\hat{n}_0$ is the direction to the point source and $\hat{n}$ is an arbitrary point on the sky. In the following three figures (Figures 46, 47, 48) we shows plots of point spread functions for both pulsar timing arrays and the LIGO Hanford-LIGO Livingston detector pair.

Example: Pulsar timing arrays

Figure 46 shows plots of point spread functions for pulsar timing arrays consisting of $N = 2$, 5, 10, 20, 25, 50 pulsars. The point source is located at the center of the maps, indicated by a black dot. The pulsar locations (indicated by white stars) were randomly-distributed on the sky, and we used equal-noise weighting for calculating the point spread function. One can see that the point spread function becomes tighter as the number of pulsars in the array increases. Figure 47 are similar plots for an actual array of $N = 20$ pulsars given in Table 6. Note that the pulsar locations are concentrated in the direction

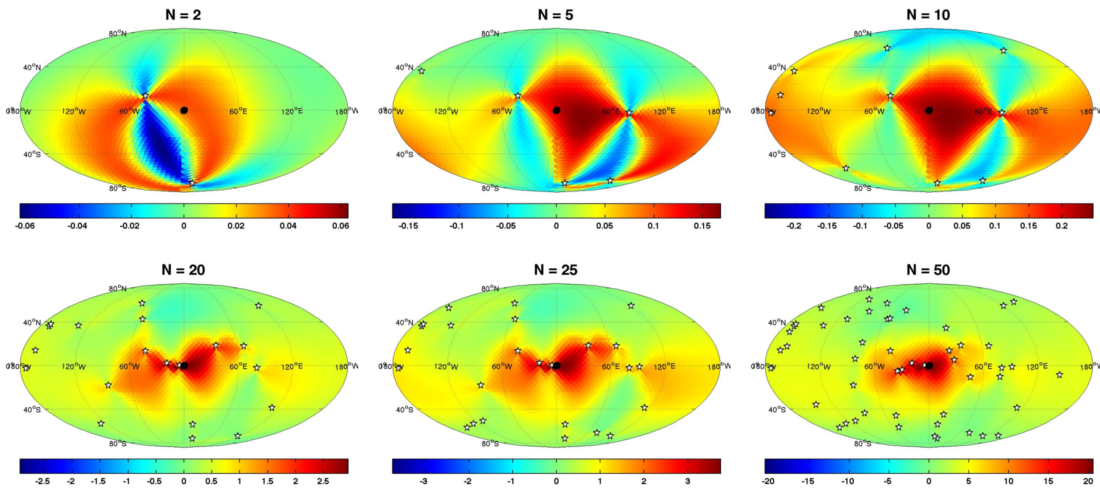


Figure 46: Point spread functions for gravitational-wave power for pulsar timing arrays consisting of $N = 2$, 5, 10, 20, 25, 50 pulsars. The point source is located at the center of the maps, $(\theta, \phi) = (90^\circ, 0^\circ)$, indicated by a black dot. The pulsar locations (indicated by white stars) are randomly placed on the sky. The point spread function becomes tighter as the number of pulsars in the array increases.

of the galactic center, $(\mathrm{ra}, \mathrm{dec}) = (-6^{\mathrm{h}}15^{\mathrm{m}}, -29^\circ)$ in equatorial coordinates. The point source is again located at the center of the maps, indicated by a black dot. The left panel shows the point spread function calculated using equal-noise weighting, while the right panel shows the point spread function calculated using *actual-noise* weighting, based on the timing noise values given in the second column of Table 6. Note that this latter plot is similar to the small-N plots in Figure 46, being dominated by pulsars with low timing noise—in this particular case, J0437-4715 and J2124-3358, having the lowest and third-lowest timing noise, dominate.

Example: Earth-based interferometers

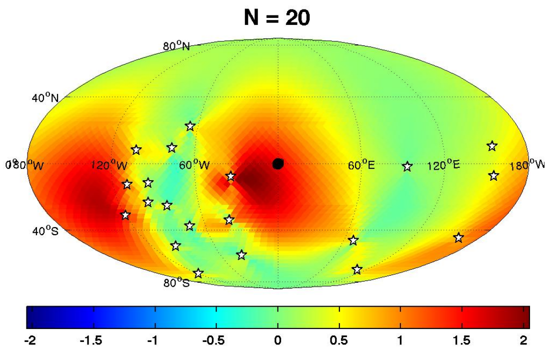


Figure 47: Point spread functions for the array of $N = 20$ pulsars listed in Table 6 for both *equal-noise* weighting (left panel) and *actual-noise* weighting (right panel), using the timing noise values in the second column of the Table. The timing noise values were rescaled by an overall factor so that the maps for the two different weighting schemes could be meaningfully compared with one another. The point source is located at the center of the maps, indicated by a black dot.

pulsar name	timing noise (μs)	pulsar name	timing noise (μs)
J0437-4715	0.14	J1730-2304	0.51
J0613-0200	2.19	J1732-5049	1.81
J0711-6830	1.04	J1744-1134	0.17
J1022+1001	0.60	J1824-2452	3.62
J1024-0719	0.35	J1909-3744	0.56
J1045-4509	3.24	J1939+2134	3.58
J1600-3053	2.67	J2124-3358	0.25
J1603-7202	1.64	J2129-5721	2.55
J1643-1224	4.86	J2145-0750	0.50
J1713+0747	0.89	B1855+0900	0.70

Table 6: Actual pulsar locations and timing noise. The pulsar name specifies its location: the first four digits is right ascension (ra) in hours and minutes (hhmm); the last four digits is declination (dec) in degrees and minutes (ddmm), with the preceding + or $-$ sign. The rms timing noise is in microsec.

In Figure 48 we plot point spread functions for gravitational-wave power for the LIGO Hanford-LIGO Livingston pair of detectors. The left hand plot is for a point source located at the center of the map, $(\theta, \phi) = (90°, 0°)$, while the right hand plot is for a point source located at $(\theta, \phi) = (60°, 0°)$ (indicated by black dots). We assumed equal white-noise power spectra for the two detectors, and we combined the contributions from 100 discrete frequencies between 0 and 100 Hz, and 100 discrete time chunks over the course of one sidereal day. The point spread functions for the two different point source locations are shaped, respectively, like a *figure-eight* with a bright region at the center of the figure-eight pattern, and a *tear drop* with a bright region near the top of the drop. These results are in agreement with [122] (see e.g., Figure 1 in that paper). Provided one combines data over a full sidereal day, the point spread function is independent of the right ascension (i.e., azimuthal) angle of the source. Readers should see [122] for more details, including a stationary phase approximation for calculating the point spread function.

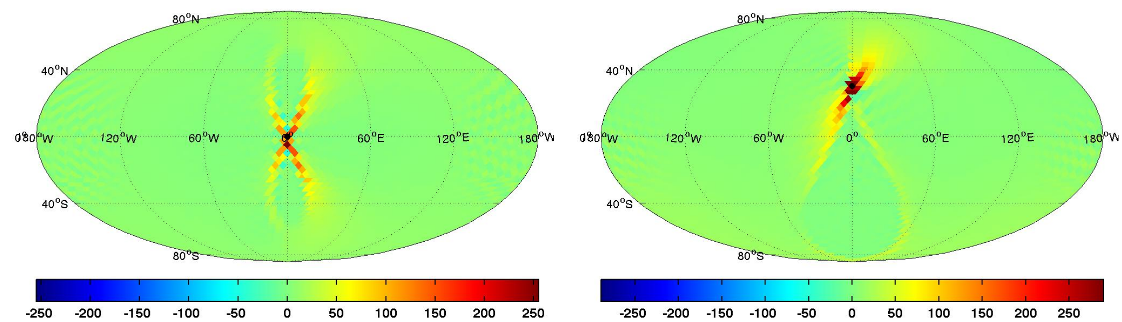


Figure 48: Point spread functions for gravitational-wave power for the LIGO Hanford-LIGO Livingston detector pair. Left panel: point source at the center of the map, $(\theta, \phi) = (90°, 0°)$. Right panel: point source at $(\theta, \phi) = (60°, 0°)$.

7.3.5 Singular value decomposition

Expression (7.33) for the maximum-likelihood estimator $\hat{\mathcal{P}}$ involves the inverse of the Fisher matrix F. But this is just a *formal* expression, as F is typically a singular matrix, requiring some sort of regularization to invert. Here we describe how *singular value decomposition* [136] can be used to 'invert' F. Since this a general procedure, we will frame our discussion in terms of an arbitrary matrix S.

Singular value decomposition factorizes an $n \times m$ matrix S into the product of three matrices:

$$S = U \Sigma V^\dagger \,, \tag{7.42}$$

where U and V are $n\times n$ and $m\times m$ unitary matrices, and Σ is an $n\times m$ rectangular matrix with (real, non-negative) singular values σ_k along its diagonal, and with zeros everywhere else. We will assume, without loss of generality, that the singular values are arranged from largest to smallest along the diagonal. We define the *pseudo-inverse* S^+ of S as

$$S^+ \equiv V \Sigma^+ U^\dagger \,, \tag{7.43}$$

where Σ^+ is obtained by taking the reciprocal of each nonzero singular value of Σ, leaving all the zeros in place, and then transposing the resulting matrix. Note that when S is a square matrix with non-zero determinant, then the pseudo-inverse S^+ is identical to the ordinary matrix inverse S^{-1}. Thus, the pseudo-inverse of a matrix generalizes the notion of ordinary inverse to non-square or singular matrices.

As a practical matter, it is important to note that if the nonzero singular values of Σ vary over several orders of magnitude, it is usually necessary to first set to zero (by hand) all nonzero singular values $\leq$ some minimum threshold value $\sigma_{\rm min}$ (e.g., 10^{-5} times that of the largest singular value). Alternatively, we can set those very small singular values equal to the threshold value $\sigma_{\rm min}$. This procedure helps to reduce the noise in the maximum-likelihood estimates, which is dominated by the modes that we are least sensitive to.

Returning to the gravitational-wave case, the above discussion means that all of the previous expressions for the inverse of the Fisher matrix, F^{-1}, should actually be written in terms of the pseudo-inverse F^+. Thus,

$$\hat{\mathcal{P}} = F^+ X \,, \tag{7.44}$$

which then implies

$$\begin{aligned} \langle\hat{\mathcal{P}}\rangle &= F^+ F\,\mathcal{P}\,, \\ \langle\hat{\mathcal{P}}\hat{\mathcal{P}}^\dagger\rangle - \langle\hat{\mathcal{P}}\rangle\langle\hat{\mathcal{P}}^\dagger\rangle &\approx F^+ \,. \end{aligned} \tag{7.45}$$

So $\hat{\mathcal{P}}$ is actually a *biased* estimator of $\mathcal{P}$ if $F^+ \neq F^{-1}$, as was discussed in [175].

Figure 49 is a plot of the singular values of typical Fisher matrices for different ground-based interferometer detector pairs (Hanford-Livingston, Hanford-Virgo, Livingston-Virgo) and a multibaseline detector network (Hanford-Livingston-Virgo). For these examples, we chose to expand the gravitational-wave power on the sky $\mathcal{P}(\hat{n})$ and the integrand of the overlap functions $\gamma_{IJ}(t; f, \hat{n})$ in terms of spherical harmonics out to $l_{\rm max} = 20$. (See Section 7.3.6 for more details about the spherical harmonic decomposition method.) This yields $(l_{\rm max} + 1)^2 = 441$ modes of gravitational-wave sky that we would like to recover. Note how the inclusion of more detectors to the network reduces the dynamic range of the singular values of F, hence making the matrix less singular without any external form of regularization.

7.3.6 Radiometer and spherical harmonic decomposition methods

The gravitational-wave radiometer [35, 34, 122] and spherical harmonic decomposition methods [175, 5] are two different ways of implementing the maximum-likelihood approach for mapping gravitational-wave power $\mathcal{P}(\hat{n})$. They differ primarily in their choice of signal model, and their approach for deconvolving the detector response from the underlying (true) distribution of power on the sky.

Gravitational-wave radiometer

The radiometer method takes as its signal model a point source characterized by a direction $\hat{n}_0$ and amplitude $\eta_{\hat{n}_0}$:

$$\mathcal{P}(\hat{n}) = \eta_{\hat{n}_0}\, \delta^2(\hat{n}, \hat{n}_0)\,. \tag{7.46}$$

It is applicable to an anisotropic gravitational-wave background dominated by a limited number of widely-separated point sources. As the number of point sources increases or if two point sources are sufficiently close to one another, the point spread function for the detector network will cause the separate signals to interfere with one another. Thus, the

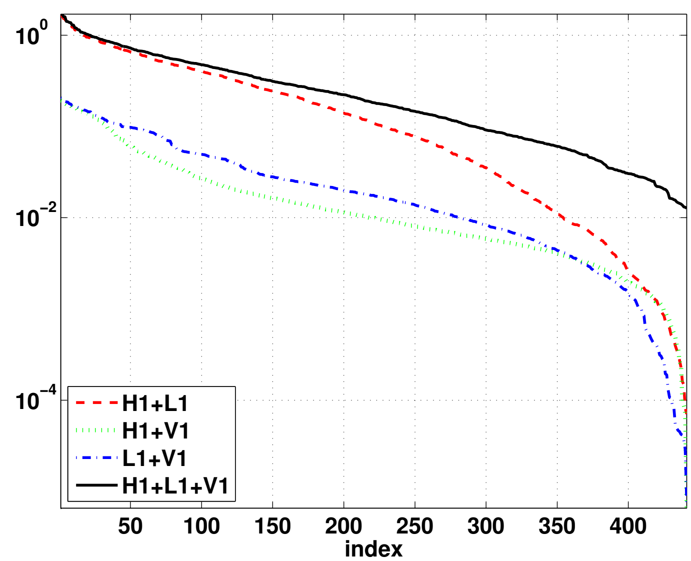


Figure 49: Singular values of typical Fisher matrices F for different ground-based interferometer detector pairs and a multibaseline detector network. For this analysis there were 441 total modes. For each individual detector pair, some of the singular values are (almost) null. The multibaseline network has fewer null modes, thus acting as a natural regularizer. Figure taken from [175].

radiometer method is not appropriate for diffuse backgrounds. Moreover, by assuming that the signal is point-like, the radiometer method ignores correlations between neigboring pixels on the sky, effectively side-stepping the deconvolution problem. Explicitly, the inverse of the Fisher matrix that appears in the maximum-likelihood estimator $\hat{\mathcal{P}} = F^{-1}X$ is replaced by the inverse of the *diagonal element* $F_{\hat{n}\hat{n}}$ to obtain an estimate of the point-source amplitude at $\hat{n}$:

$$\hat{\eta}_{\hat{n}} = (F_{\hat{n}\hat{n}})^{-1} X_{\hat{n}}\,, \tag{7.47}$$

where X is the dirty map (7.34). Thus, the radiometer method estimates the strength of point sources at different points on the sky, *ignoring* any correlations between neighboring pixels.

Note that for a single pair of detectors IJ the above estimator (7.47) is equivalent to an appropriately normalized cross-correlation statistic:

$$\hat{C}_{IJ}(t;\hat{n}) \equiv \int_{-\infty}^{\infty} df\, Q_{IJ}(t;f,\hat{n})\tilde{d}_I(t;f)\tilde{d}_J^*(t;f)\,, \tag{7.48}$$

with filter function

$$Q_{IJ}(t;f,\hat{n}) \propto \frac{\gamma_{IJ}(t;f,\hat{n})\,\bar{H}(f)}{P_{n_I}(t;f)P_{n_J}(t;f)}\,, \tag{7.49}$$

where γ_{IJ} is given by (7.8). For a network of detectors, one recovers the estimator $\hat{\eta}_{\hat{n}}$ by summing the individual-baseline statistics (7.48) over both time and distinct detector pairs, weighted by the inverse variances of the individual-baseline statistics. See e.g., [35, 34, 122] for more details.

Spherical harmonic decomposition

The spherical harmonic decomposition method is appropriate for extended anisotropic distributions on the sky, assuming a signal model for gravitational-wave power that includes spherical harmonic components up to some specified value of $l_{\rm max}$:

$$\mathcal{P}(\hat{n}) = \sum_{l=0}^{l_{\rm max}} \sum_{m=-l}^{l} \mathcal{P}_{lm} Y_{lm}(\hat{n})\,. \tag{7.50}$$

The cutoff in the expansion at $l_{\rm max}$ corresponds to an angular scale $\Delta\theta \simeq 180^\circ/l_{\rm max}$. The diffraction limit

$$\Delta\theta \simeq \frac{\lambda}{D} = \frac{c}{fD}\,, \tag{7.51}$$

where f is the maximum gravitational-wave frequency and D is the separation between a pair of detectors, sets an upper limit on the size of $l_{\rm max}$, since the detector network is not able to resolve features having smaller angular scales. For example, for the LIGO Hanford-LIGO Livingston detector pair ($D = 3000$ km) and a stochastic background having contributions out to $f \sim 1000$ Hz, we find $l_{\rm max} \lesssim 30$. Alternatively, one can use Bayesian model selection to have the data 'select' the most appropriate value of $l_{\rm max}$.

Since the spherical harmonic method targets extended distributions of gravitational-wave power on the sky, correlations between neighboring pixels or, equivalently, between different spherical harmonic components must be taken into account. This is addressed by using singular value decomposition as described in Section 7.3.5 to 'invert' the Fisher matrix. By effectively ignoring those modes that the detector network is insensitive to, we can construct the pseudo-inverse F^+ to perform the deconvolution. In terms of F^+, we have

$$\hat{\mathcal{P}}_{lm} = \sum_{l'=0}^{l_{\max}} \sum_{m'=0}^{l'} F^+_{lm,l'm'} X_{l'm'} \tag{7.52}$$

for the spherical harmonic components of the maximum-likelihood estimators $\hat{\mathcal{P}}$. The sky map constructed from the $\hat{\mathcal{P}}_{lm}$ is called a 'clean' map, since the inversion removes the detector response from the 'dirty' map X.

Figure 50 shows clean maps produced by the spherical harmonic decomposition method for a simulated anisotropic background distributed along the galactic plane [175]. The injected map is the bottom plot in the figure. (All sky maps are in equatorial coordinates.) The four maps shown in the top two rows of the figure correspond to analyses with different interferometer detector pairs (Hanford-Livingston, Hanford-Virgo, and Livingston-Virgo) and a multibaseline detector network (Hanford-Livingston-Virgo). Consistent with our findings in Figure 49, we see that the recovered map is best for the multibaseline network, whose Fisher matrix has singular values with the smallest dynamic range. For the reconstructed maps, F^+ was calculated by keeping 2/3 of all the eigenmodes (those with the largest singular values), setting the remaing singular values equal to the minimum value $\sigma_{\min}$ of the modes that were kept. For all cases, $l_{\max} = 20$. The anisotropic background was injected into simulated LIGO and Virgo detector noise (initial design sensitivity) whose power spectra are shown in Figure 51. The overall amplitude of the signal was chosen to be large enough that it was easily detectable in 1 sidereal day's worth of simulated data. For additional details see [175].

7.4 Frequentist detection statistics

As discussed in Sections 3.4 and 4.4, one can construct a frequentist detection statistic $\Lambda_{\rm ML}(d)$ by taking the ratio of the maxima of the likelihood functions for the signal-plus-noise model to the noise-only model. The logarithm,

$$\Lambda(d) \equiv 2\,\ln[\Lambda_{\rm ML}(d)]\,, \tag{7.53}$$

is the (squared) signal-to-noise ratio of the data. If we calculate this quantity for an anisotropic background $\mathcal{P}(\hat{n})$ using (7.32) for the signal-plus-noise model, we find

$$\Lambda(d) = \hat{\mathcal{P}}^\dagger F \hat{\mathcal{P}}\,, \tag{7.54}$$

where $\hat{\mathcal{P}}$ are the maximum-likelihood estimators of $\mathcal{P}$. As described in Section 3.2.1, one can use this statistic to do frequentist hypothesis testing, comparing its observed value $\Lambda_{\rm obs}$ to a threshold Λ_* to decide whether or not to claim detection of a signal.

The above detection statistic can be written in several alternative forms:

$$\Lambda(d) = \hat{\mathcal{P}}^\dagger F \hat{\mathcal{P}} = X^\dagger F^{-1} X = \frac{1}{2}\left(\hat{\mathcal{P}}^\dagger X + X^\dagger \hat{\mathcal{P}}\right)\,, \tag{7.55}$$

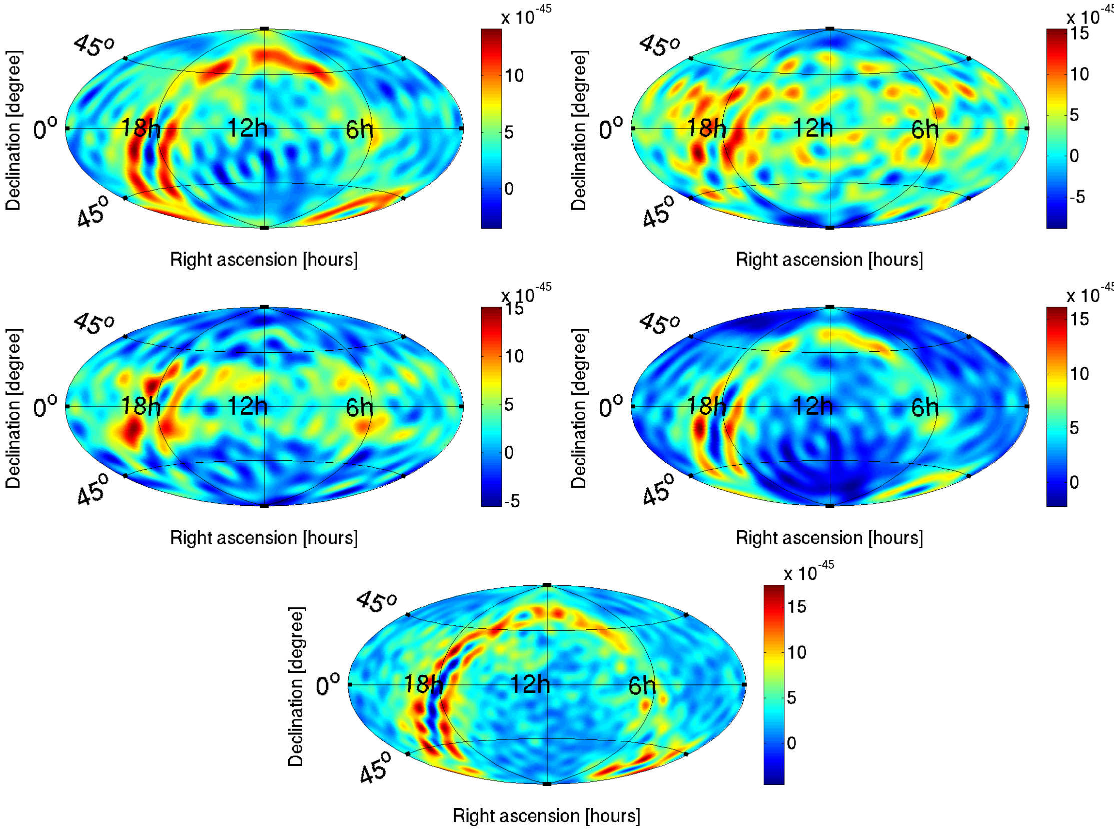


Figure 50: Results of spherical harmonic decomposition analyses performed using different detector pairs and a multibaseline detector network. The simulated anisotropic power distribution is shown in the bottom plot. Top row: Clean maps for the Hanford-Livingston and Hanford-Virgo detector pairs. Second row: Same as the top row, but for the Livingston-Virgo detector pair and for the Hanford-Livingston-Virgo multibaseline detector network. For all maps $l_{\max} = 20$. Figure taken from [175].

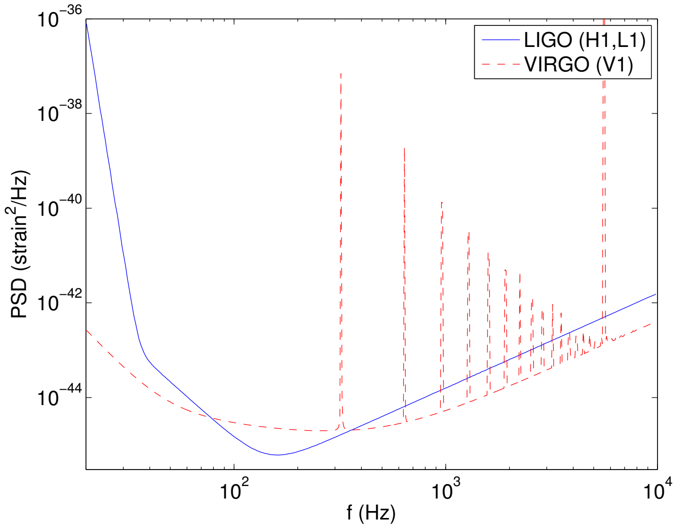


Figure 51: The power spectral densities used for the simulated detector noise for the injections described in Figure 50. Figure taken from [175].

where X is the 'dirty' map, which is related to $\hat{\mathcal{P}}$ via $\hat{\mathcal{P}} = F^{-1}X$. The last form suggests a standard matched-filter statistic:

$$\lambda(d) \equiv \frac{1}{2}\left(\bar{\mathcal{P}}^\dagger_{\rm model} X + X^\dagger \bar{\mathcal{P}}_{\rm model}\right) , \tag{7.56}$$

where $\bar{\mathcal{P}}_{\rm model}$ is an *assumed* distribution of gravitational-wave power on the sky, normalized such that

$$\bar{\mathcal{P}}^\dagger_{\rm model} F \bar{\mathcal{P}}_{\rm model} = 1 \, . \tag{7.57}$$

The above normalization is chosen so that if the true gravitational-wave background has the same spectral shape $\bar{H}(f)$ and the same angular distribution $\bar{\mathcal{P}}_{\rm model}$, then $\lambda(d)$ is an estimator of the overall amplitude of the background. In the absence of a signal, $\lambda(d)$ has zero mean and unit variance.

Such a matched-filter statistic was proposed in Appendix C of [175] and studied in detail in [164]. One nice property of this statistic is that it does not require inverting the Fisher matrix. Hence it avoids the inherent bias (7.45) and introduction of other uncertainties associated with the deconvolution process. Indeed, if we are *given* a model of the expected anisotropy, $\lambda(d)$ is the *optimal* statistic for detecting its presence. Thus, $\lambda(d)$ is especially good at detecting weak anisotropic signals. See [164] for more details.

7.5 Phase-coherent mapping

Phase-coherent mapping is an approach that constructs estimates of both the amplitude and phase of the gravitational-wave background at each point of the sky [51, 69, 142]. In some sense, it can be thought of as the "square root" of the approaches described in the previous subsections, which attempt to measure the distribution of gravitational-wave *power* $\mathcal{P}(\hat{n}) = |h_+|^2 + |h_\times^2|$. The gravitational-wave signal can be characterized in terms of either the standard polarization basis components $\{h_+(f,\hat{n}), h_\times(f,\hat{n})\}$ or the tensor spherical harmonic components $\{a^G_{(lm)}(f), a^C_{(lm)}(f)\}$. In what follows we will restrict our attention the polarization basis components, although a similar analysis can be carried out in terms of the spherical harmonic components [69].

7.5.1 Maximum-likelihood estimators and Fisher matrix

Unlike the previous approaches, which target gravitational-wave power and hence use cross-correlations (7.6) as their fundamental data product, phase-coherent mapping works directly with the data from the individual detectors. In terms of the short-term Fourier transforms defined in Section 7.1.2, we can write

$$\tilde{d}_I(t;f) = \int d^2\Omega_{\hat{n}} \sum_A R^A_I(t;f,\hat{n}) h_A(f,\hat{n}) + \tilde{n}_I(t;f)\,, \tag{7.58}$$

where I labels the different detectors, and $\tilde{n}_I(t;f)$ denotes the corresponding detector noise. Given our assumption (7.3) that the spectral and angular dependence of the background factorize with known spectral function $\bar{H}(f)$, we can rewrite the above equation as

$$\tilde{d}_I(t;f) = \int d^2\Omega_{\hat{n}}\, \bar{H}^{1/2}(f) \sum_A R^A_I(t;f,\hat{n}) h_A(\hat{n}) + \tilde{n}_I(t;f)\,, \tag{7.59}$$

so that the only unknowns are $\{h_+(\hat{n}), h_\times(\hat{n})\}$ at different locations on the sky. We will write this equation abstractly as a matrix equation

$$d = Ma + n\,, \tag{7.60}$$

where

$$M \equiv \{\bar{H}^{1/2}(f) R^A_I(t;f,\hat{n})\}\,, \qquad a \equiv \{h_A(\hat{n})\}\,. \tag{7.61}$$

The matrix multiplication corresponds to a sum over polarizations A and over directions $\hat{n}$ on the sky.

Assuming that the noise is uncorrelated across detectors, the noise covariance matrix is given by:

$$\begin{aligned} N_{Itf,I't'f'} &\equiv \langle \tilde{n}_I(t;f)\tilde{n}^*_{I'}(t';f')\rangle - \langle \tilde{n}_I(t;f)\rangle\langle \tilde{n}^*_{I'}(t';f')\rangle \\ &= \frac{\tau}{2}\delta_{II'}\delta_{tt'}\delta_{ff'} P_{n_I}(t;f)\,, \end{aligned} \tag{7.62}$$

where $P_{n_I}(t;f)$ is the one-sided power spectral density of the noise in detector I at time t. Thus, we can write down a likelihood function for the data $d \equiv \{\tilde{d}_I(t;f)\}$ given a:

$$p(d|a) \propto \exp\left[-\frac{1}{2}(d - Ma)^\dagger N^{-1}(d - Ma)\right] \tag{7.63}$$

where the multiplications inside the exponential are matrix multiplications, involving summations over detectors I, times t, and frequencies f, or summations over polarizations A

and directions $\hat{n}$ on the sky. Note that (7.63) has exactly the same form as (7.32), so the same general remarks made in Section 7.3.1 apply here as well. Namely, the maximum-likelihood estimators of a are

$$\hat{a} = F^{-1}X\,, \tag{7.64}$$

where

$$F \equiv M^{\dagger}N^{-1}M\,, \qquad X \equiv M^{\dagger}N^{-1}d\,, \tag{7.65}$$

are the Fisher matrices and 'dirty' maps for this analysis. (The definitions of M, N here are different, of course, from those in Section 7.3.1.) Explict expression for X and F are given below:

$$X \equiv X_{A\hat{n}} = \frac{2}{\tau}\sum_I\sum_t\sum_f R_I^{A*}(t;f,\hat{n})\frac{\bar{H}^{1/2}(f)}{P_{n_I}(f)}\tilde{d}_I(t;f)\,, \tag{7.66}$$

and

$$F \equiv F_{A\hat{n},A'\hat{n}'} = \frac{2}{\tau}\sum_I\sum_t\sum_f R_I^{A*}(t;f,\hat{n})\frac{\bar{H}(f)}{P_{n_I}(f)}R_I^{A'}(t;f,\hat{n}')\,. \tag{7.67}$$

Note that these expressions have an extra polarization index A, compared to the corresponding expressions for gravitational-wave power.

7.5.2 Point spread functions

The point spread function for the above analysis can now be obtained by fixing values for both A' and $\hat{n}'$, and letting A and $\hat{n}$ vary. Since there are two polarization modes (+ and $\times$), there are actually *four* different point spread functions for each direction $\hat{n}'$ on the sky:

$$\mathrm{PSF}_{AA'}(\hat{n},\hat{n}') = F_{A\hat{n},A'\hat{n}'}\,. \tag{7.68}$$

These correspond to the $A = +,\times$ responses to the $A' = +,\times$-polarized point sources located in direction $\hat{n}'$.

To illustrate the above procedure, we calculate point spread functions for phase-coherent mapping, for pulsar timing arrays consisting of $N = 1$, 2, 5, 10, 25, 50, 100 pulsars. Figure 52 show plots of these point spread functions. The pulsars are randomly distributed over the sky (indicated by white stars), and the point source is located at the center of the maps (indicated by a black dot). For simplicity, we assumed a single frequency bin, and used equal-noise weighting for calculating the point spread functions. (In addition, there is no time dependence as the directions to the pulsars are fixed on the sky.) Different rows in the figure correspond to different numbers of pulsars in the array. Different columns correspond to different choices for A and A': columns 1, 2 correspond to the $A = +,\times$ response of the pulsar timing array to an $A' = +$-polarized point source; columns 3, 4 correspond to the $A = +,\times$ response of the pulsar timing array to an $A' = \times$-polarized point source. Note that for $N = 1$, the point spread functions are proportional to either $R_I^{+}(\hat{n})$ or $R_I^{\times}(\hat{n})$ for that pulsar, producing maps similar to those shown in Figure 26. As N increases the ++ and $\times\times$ point spread functions (columns 1 and 4) become tighter around the location of the point source, which is at the center of

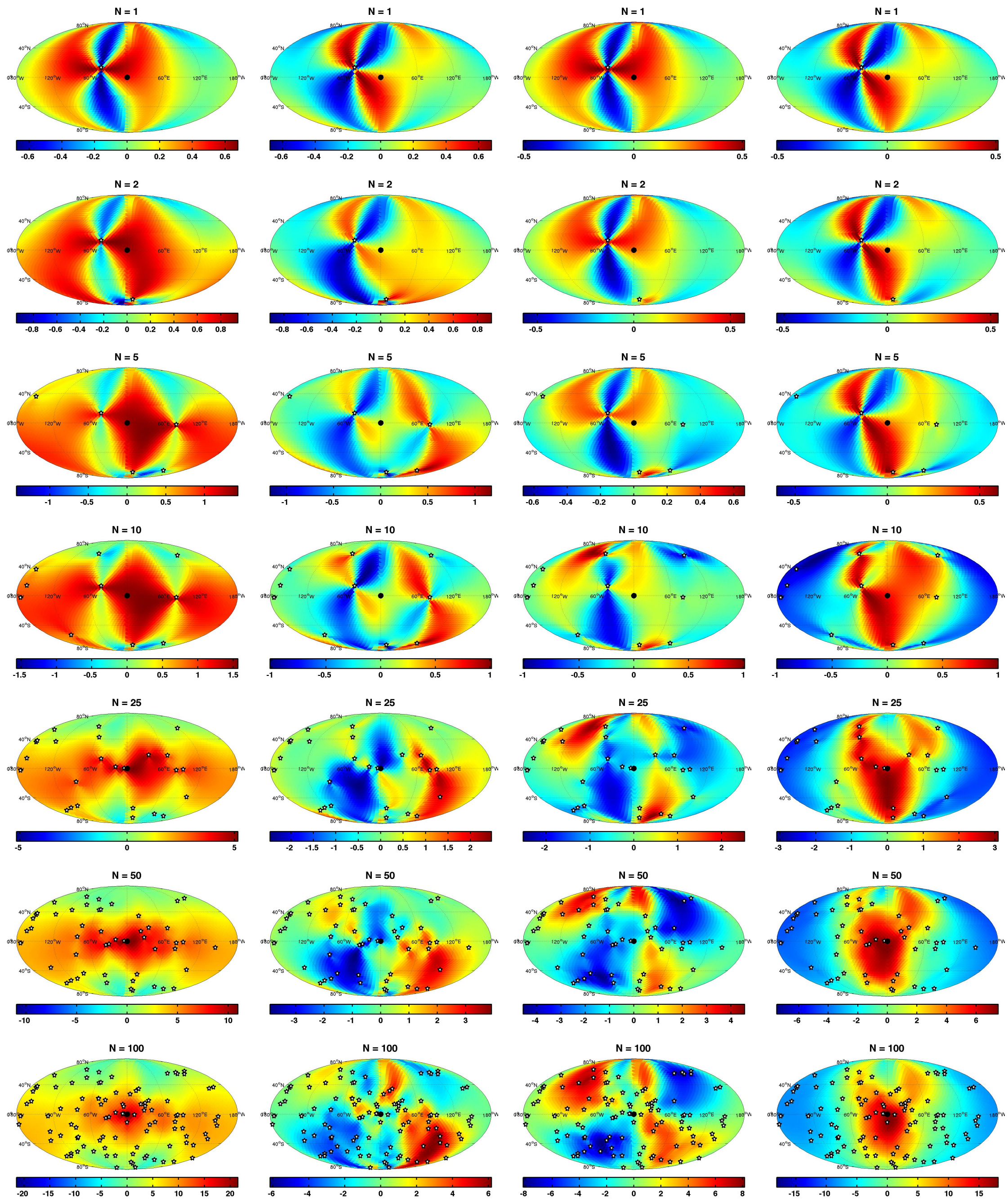


Figure 52: Point spread functions for phase-coherent mapping, for pulsar timing arrays consisting of $N = 1$, 2, 5, 10, 25, 50, 100 pulsars. The point source is located at the center of the maps, $(\theta, \phi) = (90^\circ, 0^\circ)$, indicated by a black dot. The pulsar locations (indicated by white stars) are randomly placed on the sky. Different rows correspond to different numbers of pulsars in the array. Columns 1 and 2 correspond to the $+$ and $\times$ response of the pulsar timing array to a $+$-polarized point source; columns 3 and 4 correspond to the $+$ and $\times$ response of the pulsar timing array to a $\times$-polarized point source.

the maps. But since the + and × polarizations are orthogonal, the ×+ and +× point spread functions (columns 2 and 3) have values close to zero around the location of the point source.

7.5.3 Singular value decomposition

Just as we had to deconvolve the detector response in order to obtain the estimators $\hat{\mathcal{P}}$ for gravitational-wave power, we need to do the same for the estimators $\hat{a}$ for the phase-coherent mapping approach. Although we could use singular value decomposition for the Fisher matrix F given by (7.67), we will first *whiten* the data, which lead us directly to pseudo-inverse of the whitened response matrix M, (7.61). This is the approach followed in [51, 142], and it leads to some interesting results regarding *sky-map basis vectors*, which we will describe in more detail in Section 7.5.4. An alternative approach involving the pseudo-inverse of the unwhitened response matrix is given in [69] and Appendix B of [142].

To whiten the data, we start by finding the Cholesky decomposition of the inverse noise covariance matrix $N^{-1} = LL^\dagger$, where L is a lower triangular matrix. The whitened data are then given by $\bar{d} = L^\dagger d$ (since this has unit covariance matrix), and the whitened response matrix is given by $\bar{M} = L^\dagger M$. In terms of these whitened quantities,

$$F = \bar{M}^\dagger \bar{M}\,, \qquad X = \bar{M}^\dagger \bar{d}\,, \tag{7.69}$$

implying

$$\hat{a} = F^{-1}X = (\bar{M}^\dagger \bar{M})^{-1} M^\dagger \bar{d} \equiv \bar{M}^+ \bar{d}\,. \tag{7.70}$$

The last equality is a formal expression for the pseudo-inverse $\bar{M}^+$ since $\bar{M}^\dagger \bar{M}$ is not necessarily invertible. But as shown in Section 7.3.5 it is *always* possible to define the pseudo-inverse of a matrix in terms of its singular value decomposition. Thus, given the singular value decomposition:

$$\bar{M} = U\Sigma V^\dagger\,, \tag{7.71}$$

we have

$$\bar{M}^+ = V\Sigma^+ U^\dagger\,, \tag{7.72}$$

where Σ^+ is defined by the procedure described in Section 7.3.5. Thus,

$$\hat{a} = \bar{M}^+ \bar{d} = V\Sigma^+ U^\dagger \bar{d}\,. \tag{7.73}$$

This is the expression we need to compute to calculate the maximum-likelihood estimators $\hat{a}$ for the phase-coherent mapping approach.

7.5.4 Basis skies

The singular value decomposition of $\bar{M}$ also has several nice geometrical properties. For example, from (7.73), we see that the columns of V corresponding to the non-zero singular values of Σ are *basis vectors* (which we will call *basis skies*) in terms of which $\hat{a}$ can be written as a linear combination. Similarly, if write the whitened response to the gravitational-wave background as

$$\bar{M}a = U\Sigma V^\dagger a\,, \tag{7.74}$$

then we see that the columns of U corresponding to the non-zero singular values of Σ can be interpreted as *range vectors* for the response. To be more explicit, let $u_{(k)}$ and $v_{(k)}$ denote the kth columns of U and V, and let r be the number of non-zero singular values of Σ. Then

$$\begin{aligned}\hat{a} &= \sum_{k=1}^{r} \sigma_k^{-1}(u_{(k)} \cdot \bar{d})\, v_{(k)} \,, \\ \bar{M}a &= \sum_{k=1}^{r} \sigma_k(v_{(k)} \cdot a)\, u_{(k)} \,,\end{aligned} \tag{7.75}$$

where the dot product of two vectors a and b is defined as $a \cdot b = a^\dagger b$. If we further expand $\bar{d} = \bar{M}a + \bar{n}$ in the first of these equations, then

$$\hat{a} = \sum_{k=1}^{r}(v_{(k)} \cdot a)v_{(k)} + \bar{M}^{+}\bar{n} \,. \tag{7.76}$$

This last expression involves the projection of the true gravitational-wave sky a onto the basis skies $v_{(k)}$ for only the non-zero singular values of Σ.

In Figure 53, we show plots of the real parts of the $+$ and $\times$-polarization basis skies for a pulsar timing array consisting of $N = 5$ pulsars randomly distributed on the sky. The imaginary components of the basis skies are identically zero, and hence are not shown in the figure. The basis skies are shown in decreasing size of their singular values, from top to bottom. In general, if N is the number of pulsars in the array, then the number of basis skies is $2N$ (the factor of 2 corresponding to the two polarizations, $+$ and $\times$). This means that one can extract at most $2N$ real pieces of information about the gravitational-wave background with an N-pulsar array. This is typically fewer than the number of modes of the background that we would like to recover.

7.5.5 Underdetermined reconstructions

More generally, let's consider the case where the total number of measured data points n is less than the number of modes m that we are trying to recover (so $n < m$), or where there are certain modes of the gravitational-wave background (e.g., *null skies*) that our detector network is simply insensitive to. Then, for both of these cases, the linear system of equations that we are trying to solve, $\bar{d} = \bar{M}a$, is *underdetermined*—i.e., there exist multiple solutions for a, which differ from (7.73) by terms of the form

$$a_{\rm null} = (\mathbb{1}_{m\times m} - \bar{M}^{+}\bar{M})a_{\rm arb} \,, \tag{7.77}$$

where $a_{\rm arb}$ is an *arbitrary* gravitational-wave background. (Note that $a_{\rm null}$ is an element of the *null space* of $\bar{M}$ as it maps to zero under the action of $\bar{M}$.) Our solution for $\hat{a}$ given in (7.73) sets to zero those modes that we are insensitive to. Our solution also sets to zero the variance of these modes.

In a Bayesian formulation of the problem, one needs to specify prior probability distributions for the signal parameters, in addition to specifying the likelihood function (7.63). For a mode of the background that our detector network is insensitive to, the marginalized

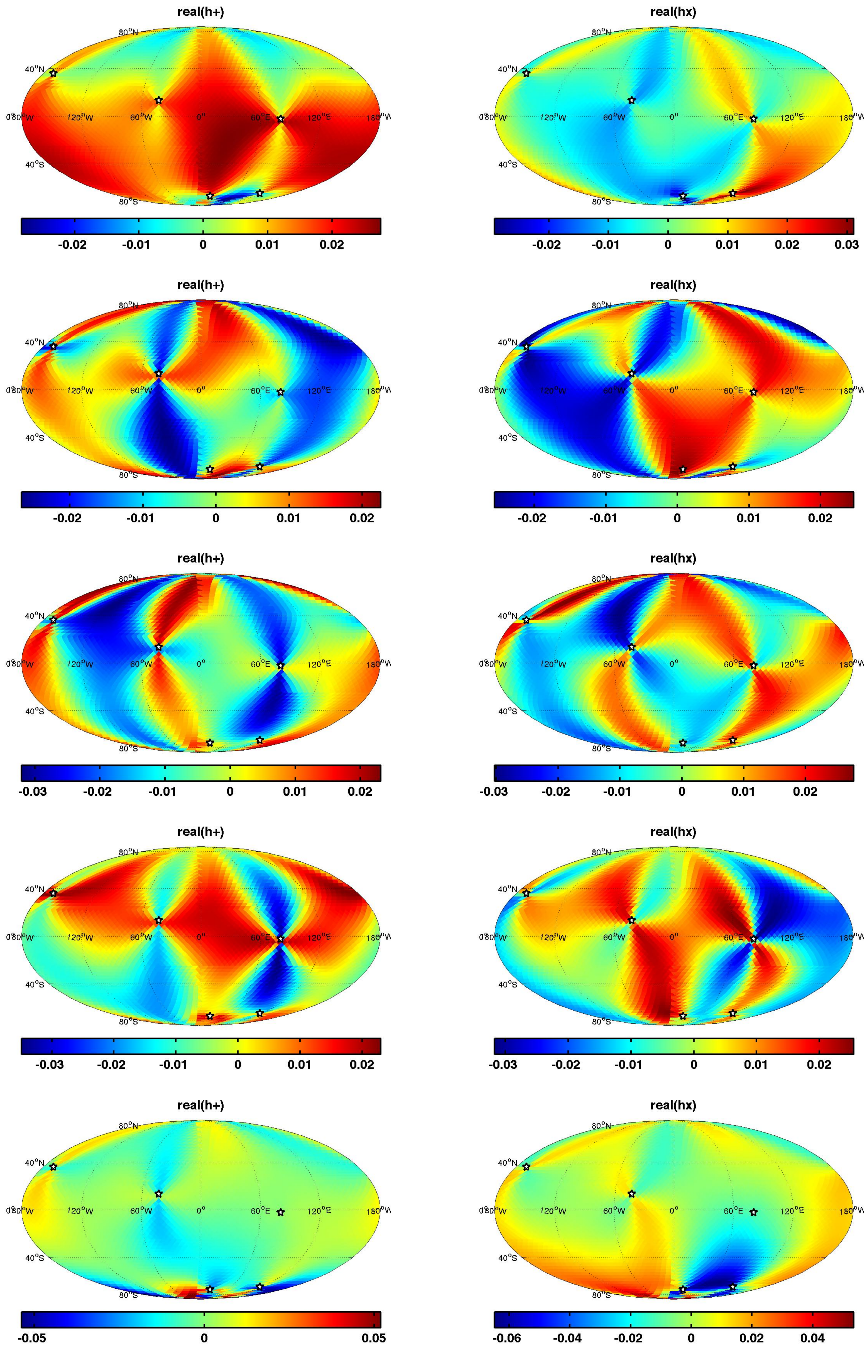


Figure 53: The real parts of the + and ×-polarization basis skies for pulsar timing array consisting of $N = 5$ pulsars randomly distributed on the sky. The imaginary components of the basis skies are identically zero. The basis skies are shown in decreasing size of their singular values, from the top of the figure to the bottom.

posterior for that mode will be the same as the prior, since the data are uniformative about this mode. This is what one would expect for a mode that is unconstrained by the data, in contrast to setting the variance equal to zero as we do with our maximum-likelihood reconstruction. Basically, our maximum-likelihood reconstruction does not attempt to say anything about the modes of the background for which we have no information.

7.5.6 Pulsar timing arrays

The phase-coherent mapping approach was first developed in the context of pulsar timing arrays [51, 69]. In [51], the analysis was done in terms of the standard polarization components $a \equiv \{h_+(f,\hat{n}), h_\times(f,\hat{n})\}$, similar to what we described above. In [69], the analysis was done in terms of the tensor spherical harmonic components $a \equiv \{a^G_{(lm)}(f), a^C_{(lm)}(f)\}$. Now recall from (5.24) that the Earth-term-only, Doppler-frequency response functions are given by

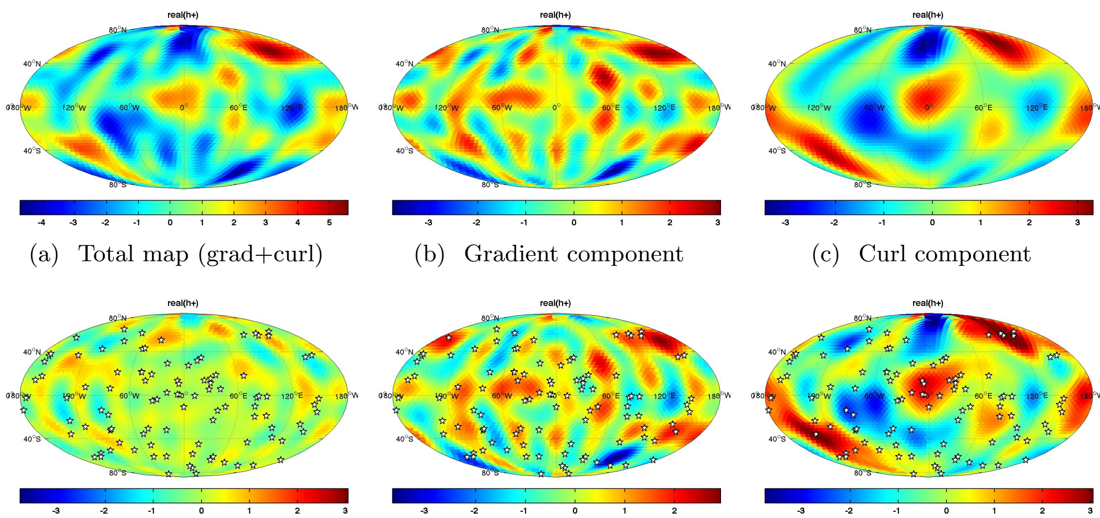
$$R^G_{(lm)}(f) = 2\pi\, {}^{(2)}\!N_l Y_{lm}(\hat{p})\,, \qquad R^C_{(lm)}(f) = 0\,, \tag{7.78}$$

where $\hat{p}$ is the direction to an arbitrary pulsar. Thus, the pulsar response to curl modes is identically zero. This means that a pulsar timing array is blind to *half* of all possible modes of a gravitational-wave background, regardless of how many pulsars there are in the array. Note that this statement is not restricted to the tensor spherical harmonic analysis; it is also true in terms of the standard $(+,\times)$ polarization components, since $a^G_{(lm)}(f)$ and $a^C_{(lm)}(f)$ are linear combinations of $h_+(f,\hat{n})$ and $h_\times(f,\hat{n})$, see (2.11). It is just that the insensivity of a pulsar timing array to half of the gravitational-wave modes is *manifest* in the gradient and curl spherical harmonic basis for which (7.78) is valid.

To explicitly demonstrate that a pulsar timing array is insensitive to the curl-component of a gravitational-wave background, Gair et al. [69] constructed maximum-likelihood estimates of a simulated background containing both gradient and curl modes. The total simulated background and its gradient and curl components are shown in the top row (panels a–c) of Figure 54. (Note that this is for a noiseless simulation so as not to confuse the lack of reconstructing the curl component with the presence of detector noise.) Panel e shows the maximum-likelihood recovered map for a pulsar timing array consisting of $N = 100$ pulsars randomly distributed on the sky. Panels d and f are residual maps obtained by subtracting the maximum-likelihood recovered map from the gradient component and the total simulated background, respectively. Note that the maximum-likelihood recovered map resembles the gradient component of the background, consistent with the fact that a pulsar timing array is insenstive to the curl component of a gravitational-wave background. The residual map for the gradient component (panel d) is much cleaner than the residual map for the total simulated background (panel f), which has angular structure that closely resembles the curl component of the background.

7.5.7 Ground-based interferometers

The phase-coherent mapping approach can also be applied to data taken by a network of ground-based interferometers [142]. Again the analysis can be performed in terms of either the standard $+$, $\times$ polarization components or the gradient and curl spherical harmonic



(a) Total map (grad+curl) (b) Gradient component (c) Curl component

(d) Gradient residual map (e) Max-likelihood recovered map (f) Total residual map

Figure 54: Mollweide projections of the real parts of $h_+(\hat{n})$ for the different components of the simulated background (panels a–c), the maximum-likelihood recovered map for a pulsar timing array consisting of $N = 100$ pulsars (panel e), and the corresponding residual maps for the grad-component (panel d) and the total simulated background (panel f). Sky maps of the imaginary part of $h_+(\hat{n})$ and the real and imaginary parts of $h_\times(\hat{n})$ are similar, and hence are not shown in this figure. Note that the maximum-likelihood recovered map most-closely resembles the gradient component of the simulated background, since a pulsar timing array is insensitive to the curl modes of a gravitational-wave background. Figure taken from [69].

components. Recall from (5.35) that

$$R^G_{(lm)}(f) = \delta_{l2}\frac{4\pi}{5}\sqrt{\frac{1}{3}}\left[Y_{2m}(\hat{u}) - Y_{2m}(\hat{v})\right] , \qquad R^C_{(lm)}(f) = 0 , \tag{7.79}$$

for a ground-based interferometer in the small-antenna limit, with its vertex at the origin, and with unit vectors $\hat{u}$, $\hat{v}$ pointing in the direction of the interferometer arms. At first, one might think that these expressions imply that a network of ground-based interferometers is also blind to the curl component of a gravitational-wave background. But (7.79) are valid only for interferometers with their vertices *at the origin* of coordinates. Since a translation mixes gradient and curl components, the response functions for an interferometer displaced from the origin by $\hat{x}_0$ are given by [142]:

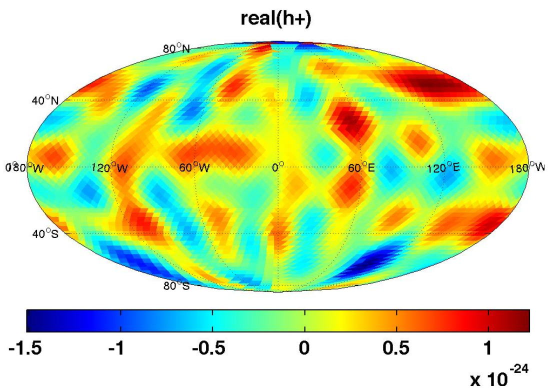
$$\begin{aligned}
R^G_{(lm)}(f) &= \sum_{m'=-2}^{2}\sum_{L=l-2}^{l+2}\sum_{M=-L}^{L} F_{m'}(\hat{u},\hat{v})4\pi(-i)^L j_L(\alpha)Y^*_{LM}(\hat{x}_0)\frac{(-1)^{m'}}{2}\left[(-1)^l + (-1)^L\right] , \\
&\quad\times\sqrt{\frac{(2\cdot 2+1)(2l+1)(2L+1)}{4\pi}}\begin{pmatrix} 2 & l & L \\ -m' & m & M\end{pmatrix}\begin{pmatrix} 2 & l & L \\ 2 & -2 & 0\end{pmatrix} , \\
R^C_{(lm)}(f) &= \sum_{m'=-2}^{2}\sum_{L=l-2}^{l+2}\sum_{M=-L}^{L} F_{m'}(\hat{u},\hat{v})4\pi(-i)^L j_L(\alpha)Y^*_{LM}(\hat{x}_0)\frac{(-1)^{m'}}{2i}\left[(-1)^l - (-1)^L\right] \\
&\quad\times\sqrt{\frac{(2\cdot 2+1)(2l+1)(2L+1)}{4\pi}}\begin{pmatrix} 2 & l & L \\ -m' & m & M\end{pmatrix}\begin{pmatrix} 2 & l & L \\ 2 & -2 & 0\end{pmatrix} ,
\end{aligned} \tag{7.80}$$

where $\alpha \equiv 2\pi f|\vec{x}_0|/c$ and $j_L(\alpha)$ are spherical Bessel functions of order L. Here

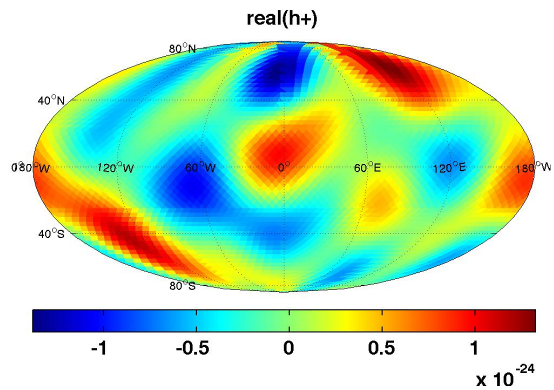
$$F_m(\hat{u},\hat{v}) \equiv \frac{4\pi}{5}\sqrt{\frac{1}{3}}\left[Y_{2m}(\hat{u}) - Y_{2m}(\hat{v})\right] , \tag{7.81}$$

is shorthand for the particular combination of spherical harmonics that enter the expression for $R^G_{(lm)}(f)$ in (7.79). The two expressions in parentheses ( ) for each response function are Wigner 3-j symbols (see for example [197, 118]). Note that the curl response is now non-zero, and both response functions depend on frequency via the quantity α, which equals 2π times the number of radiation wavelengths between the origin and the vertex of the interferometer. These expressions are valid in an arbitrary translated and rotated coordinate system, provided the angles for $\hat{u}$, $\hat{v}$, and $\hat{x}_0$ are calculated in the rotated frame.

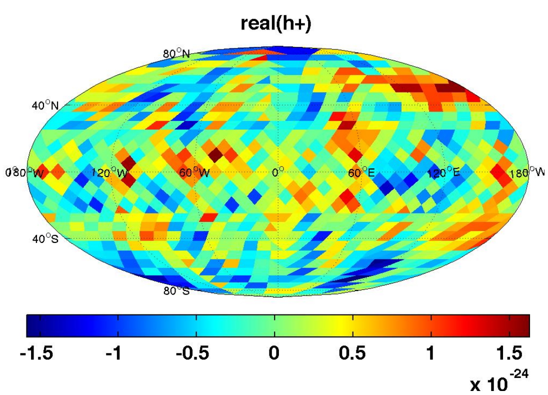
Thus, the spatial separation of a network of ground-based interferometers, or of a single interferometer at different times during its daily rotational and yearly orbital motion around the Sun (Section 5.5.3), allows for recovery of both the gradient *and* curl components of a gravitational-wave background. This is in contrast to a pulsar timing array, which is insensitive to the curl component, because one vertex of all the pulsar baselines are 'pinned' to the solar system barycenter. To illustrate this difference, we show in Figure 55, maximum-likelihood recovered sky maps for simulated grad-only and curl-only anistropic backgrounds injected into noise for a 3-detector network of ground-based interferometers (Hanford-Livingston-Virgo). The grad-only and curl-only backgrounds are the same as those used for the simulated maps in Figure 54. In contrast to the recovered maps shown in that figure for the pulsar timing array, the maximum-likelihood maps (bottom row) for the network of ground-based interferometers reproduce the general angular structure of both the grad-only *and* curl-only injected maps (shown in the top row). (The noise for these injections degrades the recovery compared to the noiseless injections in Figure 54.) See [142] for more details and related simulations.



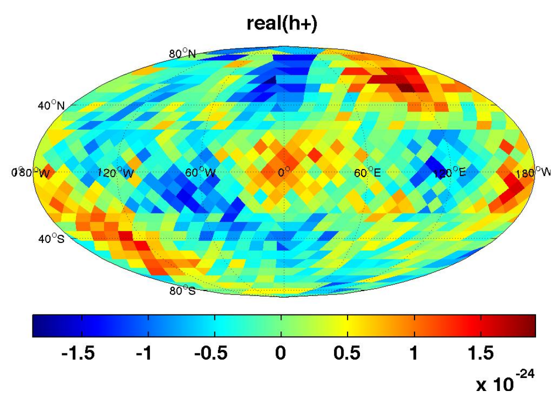
(a) Injected grad-only map



(b) Injected curl-only map



(c) Max-likelihood recovered map



(d) Max-likelihood recovered map

Figure 55: Mollweide projections of the real parts of $h_+(\hat{n})$ for grad-only and curl-only anisotropic backgrounds injected into noise and analysed using a 3-detector network of ground-based laser interferometers (Hanford-Livingston-Virgo). The injected maps are shown in the top row; the maximum-likelihood recovered maps are shown in the second row. Sky maps of the imaginary part of $h_+(\hat{n})$ and the real and imaginary parts of $h_\times(\hat{n})$ are similar for both the injections and the recovered maps, and hence are not shown in the figure. Note that a network of ground-based interferometers is capable of recovering both the gradient and curl components of a gravitational-wave background, in contrast to a pulsar timing array (compare with Figure 54). Figure taken from [142].

8 Searches for other types of backgrounds / signals

> No idea is so outlandish that it should not be considered with a searching but at the same time a steady eye. *Winston Churchill*

Since stochastic gravitational-wave backgrounds come in many different "flavors", one needs additional search methods that go beyond the standard "vanilla" cross-correlation search for a Gaussian-stationary, unpolarized, isotropic signal (Sections 4 and 5) to extract the relevant information from the more exotic backgrounds. In Section 7, we discussed how to search for *anisotropic* signals, which are stronger coming from certain directions on the sky than from others. In this section, we discuss search methods for non-Gaussian signals (Section 8.1), circularly polarized backgrounds (Section 8.2), and additional polarization modes predicted by alternative (non-general-relativity) metric theories of gravity (Sections 8.3, 8.4, 8.5). In Section 8.6, we also briefly mention searches for other types of gravitational-wave signals, which are not really stochastic backgrounds, but nonetheless can be searched for using the basic idea of cross-correlation, which we developed in Section 4. The majority of the search methods that we will describe here have been implemented "across the band"—i.e., for ground-based interferometers, space-based interferometers, and pulsar timing arrays. For these methods, we will highlight any significant differences in the implementations for the different detectors if there are any.

Of course, we do not have enough time or space in this section to do justice for all of these methods. As such, readers are strongly encouraged to read the original papers for more details. For non-Gaussian backgrounds, see [58, 153, 174, 116, 44]; for circular polarization, see [155, 156, 101]; for polarization modes in alternative theories of gravity, see [109, 130, 41, 71]; and for the othe types of signals, see [179, 117].

8.1 non-Gaussian backgrounds

In Section 2.1, we asked the question "when is a gravitational-wave signal stochastic" to highlight the practical distinction between searches for deterministic and stochastic signals. From an operational perspective, a signal is stochastic if it is best searched for using a stochastic signal model (i.e., one defined in terms of probability distributions), even if the signal is *intrinsically* deterministic, e.g., a superposition of sinusoids. This turns out to be the case if the signals are: (i) *sufficiently weak* that they are individually unresolvable in a single detector, and hence can only be detected by integrating their correlated contribution across multiple detectors over an extended period of time, or (ii) they are *sufficiently numerous* that they overlap in time-frequency space, again making them individually unresolvable, but producing a *confusion noise* that can be detected by cross-correlation methods. If the rate of signals is large enough, the confusion noise will be Gaussian thanks to the central limit theorem. But if the rate or duty-cycle is small, then the resulting stochastic signal will be non-Gaussian and "popcorn-like", as we discussed in Section 1.1. This is the type of signal that we expect from the population of binary black holes that produced GW150914 and GW151226; and it is the type of signal that we will focus on in the following few subsections.

Figure 56 illustrates the above statements in the context of a simple toy-model signal consisting of simulated sine-Gaussian bursts (each with a width $\sigma_t = 1$ s) having different rates or duty cycles. The left two panels correspond to the case where there is 1 burst every 10 seconds (on average). The probability distribution of the signal samples h (estimated by the histogram in the lower-left-hand panel) is far from Gaussian for this case. The right two panels correspond to 100 bursts every second (on average), for which the probability distribution is approximately Gaussian-distributed, as expected from the central-limit theorem.

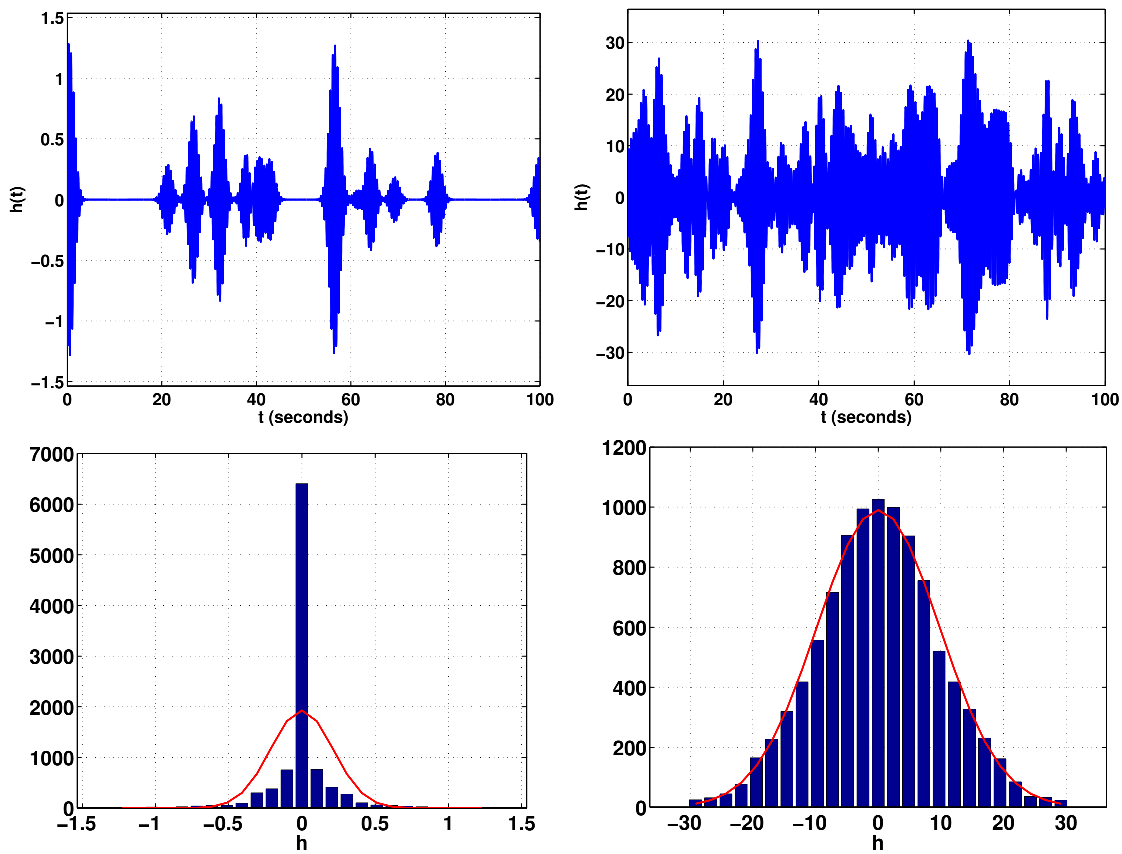


Figure 56: Simulated toy-model signals and histograms for different duty cycles. The left two panels correspond to 1 burst every 10 seconds (on average); the right two panels correspond to 100 bursts every second (on average). The red curves in the bottom two panels show the best-fit Gaussian distributions to the data. Similar to Figure 1 from [174].

8.1.1 non-Gaussian search methods – overview

There are basically two different approaches that one can take to search for non-Gaussian stochastic signals: (i) The first is to incorporate the non-Gaussianity of the signal into the likelihood function by marginalizing over the appropriate signal model (Section 8.1.2).

Then given the likelihood, one can construct frequentist detection statistics and estimators from the maximum-likelihood ratio (3.19), or do Bayesian model selection in the usual way (Section 3). (ii) The second approach is to construct specific frequentist statistics that targets the higher-order moments of the non-Gaussian distribution, and then use these statistics to do standard frequentist hypothesis testing and parameter estimation. This approach is most simply cast in terms of the *skewness* and (excess) *kurtosis* of the distribution, which are the third and fourth-order *cumulants*, defined as follows: If X is a random variable with probability distribution $p_X(x)$, then the *moments* are defined by (Appendix B):

$$\mu_n \equiv \langle X^n \rangle = \int dx\, x^n p_X(x)\,, \tag{8.1}$$

and the *cumulants* by

$$\begin{aligned}
c_1 &= \mu_1\,,\\
c_2 &= \mu_2 - \mu_1^2\,,\\
c_3 &= \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3\,,\\
c_4 &= \mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4\,,\\
&\vdots
\end{aligned} \tag{8.2}$$

Note that c_1 and c_2 are just the *mean* μ and *variance* σ^2 of the distribution. For a Gaussian distribution, $c_3 = 0, c_4 = 0, \cdots$. For a distribution with zero mean, the above formulas simplify to $c_1 = 0$, $c_2 = \mu_2$, $c_3 = \mu_3$, and $c_4 = \mu_4 - 3\mu_2^2$. The higher-order-moment approach requires 3rd or 4th-order correlation measurements (Section 8.1.5).

8.1.2 Likelihood function approach for non-Gaussian backgrounds

Fundamentally, searching for non-Gaussian stochastic signals is no different than searching for a Gaussian stochastic signal. In both cases one must: (i) specify a signal model, (ii) incorporate that signal model into a likelihood function or frequentist detection statistic/estimator, and (iii) then analyze the data to determine how likely it is that a signal is present. Its the choice of signal model, of course, that determines what type of signal is being searched for.

The signal model is incorporated into the likelihood via marginalization over the signal samples as discussed in Section 3.6.2. Assuming Gaussian-stationary noise[19] with covariance matrix C_n, the probability of observing data d in a network of detectors given signal model $\bar{h}$ is (3.49):

$$p(d|\bar{h}, C_n) = \frac{1}{\sqrt{\det(2\pi C_n)}} e^{-\frac{1}{2} r_{Ii} (C_n^{-1})_{Ii,Jj} r_{Jj}}\,, \tag{8.3}$$

where

$$r_{Ii} \equiv d_{Ii} - \bar{h}_{Ii} \tag{8.4}$$

[19]What to do when the noise is non-stationary or non-Gaussian is discussed in Sections 9.1 and 9.2.

are the residuals in detector I. (The subscript i labels either a time or frequency sample for the analysis, whichever is being used.) Since one is often not interested in the particular values of $\bar{h}$, but rather the values of the parameters $\vec{\theta}_h$ that describe the signal, one marginalizes over $\bar{h}$:

$$p(d|\vec{\theta}_h, \vec{\theta}_n) = \int d\bar{h}\, p(d|\bar{h}, C_n) p(\bar{h}|\vec{\theta}_h)\,. \tag{8.5}$$

This yields a likelihood function that depends on the signal and noise parameters $\vec{\theta}_h$, $\vec{\theta}_n \equiv C_n$. It is this likelihood function that we then use for our statistical analysis.

Several different signal priors, which have been proposed in the literature, are given below. For simplicity, we will consider the case where the detectors are colocated and coaligned, and have isotropic antenna patterns, so that the contribution from the signal is the same in each detector, and is independent of direction on the sky. For real analyses, these simplifications will need to be dropped, as is done e.g., in [174].

Gaussian signal prior:

$$p(\bar{h}|S_h) = \frac{1}{(2\pi S_h)^{N/2}}\, e^{-\frac{1}{2S_h}\sum_{i=1}^N \bar{h}_i^2}\,. \tag{8.6}$$

This is the standard prior that one uses for describing a Gaussian-stochastic signal, and leads to the usual Gaussian-stochastic cross-correlation detection statistic (Section 4.4).

Drasco and Flanagan [58] non-Gaussian signal prior:

$$p(\bar{h}|\xi, \alpha) = \prod_{i=1}^N \left[\xi\, \frac{1}{\sqrt{2\pi\alpha^2}}\, e^{-\bar{h}_i^2/2\alpha^2} + (1-\xi)\,\delta(\bar{h}_i)\right]\,. \tag{8.7}$$

This prior corresponds to Gaussian bursts occuring with probability $0 \le \xi \le 1$ and with root-mean-square (rms) amplitude α.

Mixture-Gaussian signal prior:

$$p(\bar{h}|\xi, \alpha, \beta) = \prod_{i=1}^N \left[\xi\, \frac{1}{\sqrt{2\pi\alpha^2}}\, e^{-\bar{h}_i^2/2\alpha^2} + (1-\xi)\, \frac{1}{\sqrt{2\pi\beta^2}}\, e^{-\bar{h}_i^2/2\beta^2}\right]\,. \tag{8.8}$$

The mixture-Gaussian signal prior is a non-Gaussian distribution, which reduces to the Gaussian signal prior in the limit $\xi \to 1$. It reduces to the Drasco and Flanagan signal prior in the limit $\beta \to 0$.

Martellini and Regimbau [116] non-Gaussian signal prior:

$$p(\bar{h}|\xi, \alpha) = \prod_{i=1}^N \left[\xi\, p_{\rm NG}(\bar{h}_i) + (1-\xi)\,\delta(\bar{h}_i)\right]\,, \tag{8.9}$$

where

$$p_{\rm NG}(\bar{h}_i) = \frac{1}{\sqrt{2\pi\alpha^2}}\, e^{-\bar{h}_i^2/2\alpha^2} \left[1 + \frac{c_3}{6\alpha^3} H_3\left(\frac{\bar{h}_i}{\alpha}\right) + \frac{c_4}{24\alpha^4} H_4\left(\frac{\bar{h}_i}{\alpha}\right) + \frac{c_3^2}{72\alpha^6} H_6\left(\frac{\bar{h}_i}{\alpha}\right)\right] \tag{8.10}$$

is the 4th-order Edgeworth expansion [116] of a non-Gaussian distribution with third and fourth-order cumulants c_3 and c_4. ($H_n(x)$ denotes a Hermite polynomial of order n.) The Edgeworth expansion is referenced off a Gaussian probability distribution, and is thus said to be a *semi-parametric* representation of a non-Gaussian distribution. This prior reduces to the Drasco and Flanagan signal prior when $c_3 = 0$, $c_4 = 0$.

Multi-sinusoid signal prior:

$$\begin{aligned} p(\bar{h}|\vec{\theta}_h) &= \delta\left(\bar{h} - \bar{h}(\vec{\theta}_h)\right)\,, \\ \bar{h}_i(\vec{\theta}_h) &= \sum_{I=1}^{M} A_I \cos(2\pi f_I t_i - \varphi_I)\,. \end{aligned} \tag{8.11}$$

This is a *deterministic* signal prior, corresponding to the superposition of M sinusoids with unknown amplitudes, frequencies and phases, $\vec{\theta}_h = \{A_I, f_I, \varphi_I | I = 1, 2, \cdots, M\}$. This was one of the signal models used in [44] to investigate the question of when is a signal stochastic.

Superposition of finite-duration deterministic signals:

$$\begin{aligned} p(\bar{h}|\vec{\theta}_h) &= \delta\left(\bar{h} - \bar{h}(\vec{\theta}_h)\right)\,, \\ \bar{h}_i(\vec{\theta}_h) &= \sum_{I=1}^{M} A_I \mathcal{T}(t_i - t_I|\vec{\theta}_{\mathcal{T}})\,. \end{aligned} \tag{8.12}$$

Here, $\mathcal{T}(t|\vec{\theta}_{\mathcal{T}})$ is a normalized waveform (template) for some deterministic signal (e.g., a chirp from an inspiralling binary, a sine-Gaussian burst, a ringdown signal, $\cdots$) described by parameters $\vec{\theta}_{\mathcal{T}}$ (e.g., chirp mass, correlation time, frequency, $\cdots$). A_I is the amplitude of the Ith signal and t_I is its arrival time. Note that these signal waveforms can be *extended* in time, having a characteristic duration τ. Thus, this signal model is intermediate between the single-sample burst and multi-sinusoid signal models.

Generic likelihood for unresolvable signals:

In [174], Thrane writes down a generic likelihood function for a non-Gaussian background formed from the superposition of signals which are individually unresolvable in a single detector. The likelihood function:

$$p(\hat{\rho}|\xi, \vec{\theta}_h, \vec{\theta}_n) = \prod_i \left[\xi\, S(\hat{\rho}_i|\vec{\theta}_h) + (1-\xi)\, B(\hat{\rho}_i|\vec{\theta}_n)\right] \tag{8.13}$$

is defined for a pair of detectors I, J, and takes as its fundamental data vector estimates of the signal-to-noise ratio of the cross-correlated power in the two detectors:

$$\hat{\rho}_i \equiv \hat{\rho}(t; f) = \sqrt{\tau \delta f}\, \frac{\hat{C}_{IJ}(t; f)}{\sqrt{P_{n_I}(t; f) P_{n_J}(t; f)}}\,, \tag{8.14}$$

where

$$\hat{C}_{IJ}(t; f) \equiv \frac{2}{\tau} \tilde{d}_I(t; f) \tilde{d}_J^*(t; f)\,. \tag{8.15}$$

Here τ is the duration of the short-term Fourier transforms and δf is the frequency resolution. (Note that δf can be greater than $1/\tau$ if one averages together neighboring frequency bins.) The product over i is over time-frequency pixels tf. The functions S and B are probability distributions for $\hat{\rho}_i$ for the signal and noise models, respectively. These distributions are generic in the sense that they are to be estimated using Monte Carlo simulations with injected signals for the signal model S, and via time-slides on real data for the noise model B. They need not be Gaussian for either the signal or the detector noise. The vectors $\vec{\theta}_h$ and $\vec{\theta}_n$ denote parameters specific to the signal and noise models. Although the above likelihood function was not obtained by explicitly marginalizing over $\bar{h}$, mathematically there is some signal prior and noise model which yields this likelihood upon marginalization.

8.1.3 Frequentist detection statistic for non-Gaussian backgrounds

As discussed in Section 3.4, given likelihood functions for the signal-plus-noise and noise-only models, we can construct a frequentist detection statistic from either the maximum-likelihood ratio $\Lambda_{\rm ML}(d)$ given by (3.19), or twice its logarithm, $\Lambda(d) \equiv 2\ln(\Lambda_{\rm ML}(d))$, which has the interpretation of being the squared signal-to-noise ratio of the data. For a white Gaussian stochastic signal in white Gaussian detector noise (assuming a pair of colocated and coaligned detectors), we showed in Section 4.4:

$$\Lambda^{\rm G}_{\rm ML}(d) = \left[1 - \frac{\hat{S}_h^2}{\hat{S}_1\hat{S}_2}\right]^{-N/2}, \qquad \Lambda^{\rm G}(d) \approx \frac{\hat{S}_h^2}{\hat{S}_{n_1}\hat{S}_{n_2}/N}, \tag{8.16}$$

where N is the number of samples, and where the last approximate equality assumes that the gravitational-wave signal is weak compared to the detector noise. We have added the superscript G to indicate that this is for a Gaussian-stochastic signal model.

We can do the exactly the same calculations, making the same assumptions, for the likelihood functions constructed from *any* of the non-Gaussian signal priors given above (in Section 8.1.2). These calculations have already been done for the Drasco-Flanagan and Martellini-Regimbau signal priors [58, 116]. The expressions that they find for the maximum-likelihood ratios $\Lambda^{\rm NG}_{\rm ML}(d)$ for their non-Gaussian signal models are rather long and not particularly informative, so we do not bother to write them down here (interested readers should see (1.8) in [58] and the last equation in [116].) The values of the parameters that maximize the likelihood ratio are estimators of ξ, α, S_{n_1}, S_{n_2} for the Drasco and Flanagan signal model, and estimators of ξ, α, c_3, c_4, S_{n_1}, S_{n_2} for the Martellini and Regimbau signal model.

To illustrate the performance of a non-Gaussian detection statististic, we plot in Figure 57 the mimimum value of $\Omega_{\rm gw}$ (S_h in the notation above) necessary for detection as a function of the duty cycle ξ. (The signal becomes Gaussian as $\xi \to 1$.) The solid line is the theoretical prediction for the Drasco and Flanagan non-Gaussian maximum-likelihood statistic, while the dashed line is the theoretical prediction for the standard Gaussian-stochastic cross-correlation statistic. The dotted line is the theoretical prediction for a single-detector *burst* statistic, which is just the maximum of the absolute value of the data samples in e.g., detector 1: $\Lambda^{\rm B}(d) = \max_i |d_{1i}|$. The false alarm and false dismissal probabilities were both chosen to equal 0.01 for this calculation. From the figure one sees that for $\xi \gtrsim 10^{-3}$, the Gaussian-stochastic cross-correlation statistic performs best. For smaller values of ξ, the non-Gaussian statistic is better. In particular, for $\xi \sim 10^{-4}$. there is a factor of ~ 2 improvement in the minimum detectable signal amplitude if one uses the non-Gaussian maximum-likelihood detection statistic.

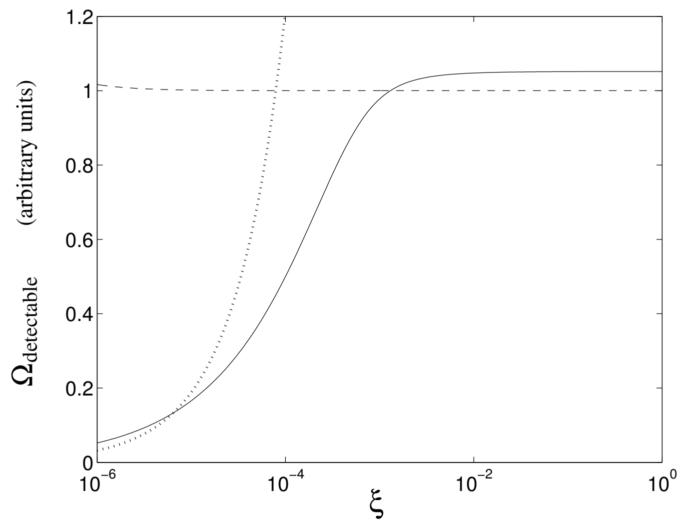


Figure 57: The mimimum detectable value of $\Omega_{\rm gw}$ as a function of the duty cycle ξ. The solid line is the theoretical prediction for the Drasco and Flanagan non-Gaussian maximum-likelihood statistic; the dashed line is for the standard Gaussian-stochastic cross-correlation statistic; and the dotted line is for a single-detector burst statistic. The number of data points used was $N = 10^9$, and the false alarm and false dismissal probabilities were both chosen to equal 0.01. (Figure taken from [58].)

Figure 58 is taken from [174] and shows posterior distributions for the duty cycle ξ calculated for Monte Carlo simulations corresponding to pure background $\xi = 0$ (dash-dot blue), pure signal $\xi = 1$ (solid red), and an even mixture $\xi = 0.5$ (dashed green). These curves illustrate that the formalism in [174] can provide estimates of the duty cycle ξ of the non-Gaussian background. See [174] for more details.

8.1.4 Bayesian model selection

As an alternative to using frequentist detection statistics and estimators to search for potentially non-Gaussian signals, one can use Bayesian model selection to compare the noise-only model $\mathcal{M}_0$ to different signal-plus-noise models $\mathcal{M}_1, \mathcal{M}_2, \cdots$. This is a general procedure for Bayesian inference, which was discussed in Section 3.3.3. As shown there, the posterior odds ratio between two different models $\mathcal{M}_\alpha$ and $\mathcal{M}_\beta$ can be written as

$$\mathcal{O}_{\alpha\beta}(d) = \frac{p(\mathcal{M}_\alpha|d)}{p(\mathcal{M}_\beta|d)} = \frac{p(\mathcal{M}_\alpha)}{p(\mathcal{M}_\beta)}\,\frac{p(d|\mathcal{M}_\alpha)}{p(d|\mathcal{M}_\beta)}\,, \tag{8.17}$$

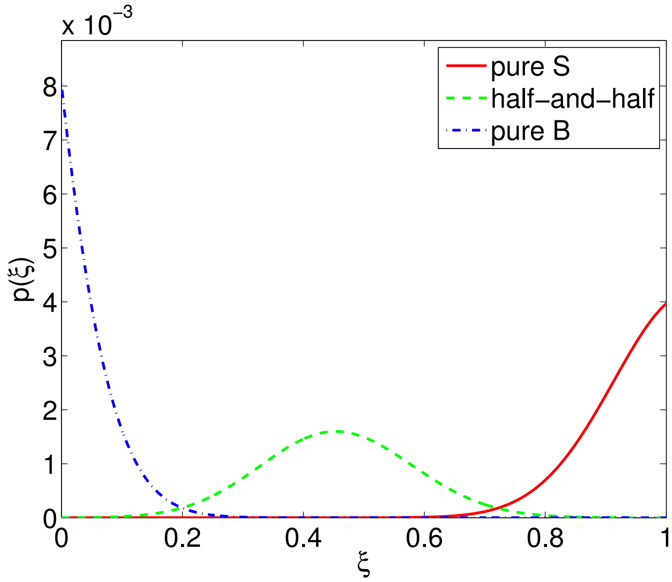


Figure 58: Posterior distributions for the duty cycle ξ calculated for Monte Carlo simulations having $\xi = 0$ (dash-dot blue), $\xi = 1$ (solid red), and $\xi = 0.5$ (dashed green). (Figure taken from [174].)

where the first ratio on the right-hand side is the *prior* odds for the two models, while the second term is the *Bayes factor*:

$$\mathcal{B}_{\alpha\beta}(d) = \frac{p(d|\mathcal{M}_\alpha)}{p(d|\mathcal{M}_\beta)}\,, \tag{8.18}$$

which is a ratio of model evidences:

$$p(d|\mathcal{M}_\alpha) = \int p(d|\vec{\theta}_\alpha, \mathcal{M}_\alpha)p(\vec{\theta}_\alpha|\mathcal{M}_\alpha)\,d\vec{\theta}_\alpha\,, \tag{8.19}$$

and similarly for $p(d|\mathcal{M}_\beta)$. If one assumes equal prior odds, then the posterior odds ratio is just the Bayes factor, and we can use its value to rule in favor of one model or another (see Table 3).

The idea of using Bayesian model selection in the context of searches for non-Gaussian stochastic backgrounds was proposed by us in [44]. We considered a simple toy-problem consisting of simulated data in two colocated and coaligned detectors, having uncorrelated white Gaussian detector noise plus a gravitational-wave signal formed from the superposition of sinusoids having amplitudes drawn from an astrophysical population of sources. Such a signal is effectively the *frequency-domain version* of the short-duration time-domain bursts discussed in the previous subsections. Five different models were considered:

- $\mathcal{M}_0$: noise-only model, consisting of uncorrelated white Gaussian noise in two detectors with unknown variances σ_1^2, σ_2^2.

- $\mathcal{M}_1$: noise plus the Gaussian-stochastic signal model defined by (8.6).
- $\mathcal{M}_2$: noise plus the mixture-Gaussian stochastic signal model defined by (8.8).
- $\mathcal{M}_3$: noise plus the deterministic multisinusoid model defined by (8.11).
- $\mathcal{M}_4$: noise plus the deterministic multisinusoid signal model plus the Gaussian-stochastic signal model. This is a *hybrid* signal model that allows for both stochastic and determistic components for the signal.

Simulated data were generated by coadding sinusoidal signals with amplitudes drawn from an astrophysical model [150], and phases and frequencies drawn uniformly across the range spanned by the data. Gaussian-distributed white noise for the two detectors were then added to the signal data. The amplitude of the signals were scaled so as to produce a specified matched filter signal-to-noise ratio per frequency bin. Markov chain Monte Carlo analyses were run to compare the noise-only model $\mathcal{M}_0$ to each of the four signal-plus-noise models $\mathcal{M}_1, \cdots, \mathcal{M}_4$. Quantile intervals for the Bayes factors were estimated from 256 independent realizations of the simulated data for each set of parameter values. These intervals capture the fluctuation in the Bayes factors that come from *different* realizations of the data; they are not uncertainties in the Bayes factors associated with different Monte Carlo simulations for a *single* realization, which were $\lesssim 10\%$.

Figure 59 is a representative plot taken from [44], comparing the different models. The left panel shows the Bayes factors for the four different signal-plus-noise models relative to the noise-only model plotted as a function of the average number of sources per bin. The right panel shows the fraction of time that the different models had the largest Bayes factor for the different simulations. The total number of bins was set to 32 for these simulations and the SNR per bin was fixed at 2. From these and other similar plots in [44], one can draw the general conclusion that deterministic models are generally favored for small source densities, a non-Gaussian stochastic model is preferred for intermediate source densities, and a Gaussian-stochastic model is preferred for large source densities. Given the large fluctuations in the Bayes factors associated with different signal realizations, the boundaries between these three regimes is rather fuzzy. The hybrid model, which has a deterministic component for the bright signals and a Gaussian-stochastic component for the remaining confusion background, is the best model for the majority of cases.

8.1.5 Fourth-order correlation approach for non-Gaussian backgrounds

In this section, we briefly describe a fourth-order correlation approach for detecting non-Gaussian stochastic signals, originally proposed in [153]. The key idea is that by forming a particular combination of data from 4 detectors (the *excess kurtosis*), one can separate the non-Gaussian contribution to the background from any Gaussian-distributed component. This approach requires that the noise in the four detectors be uncorrelated with one another, but it does not require that the noise be Gaussian. Here we sketch out the calculation for 4 colocated and coaligned detectors, which we will assume have isotropic antenna patterns, so that the contribution from the gravitational-wave signal is the same in each detector, and is independent of direction on the sky. These simplifying assumptions

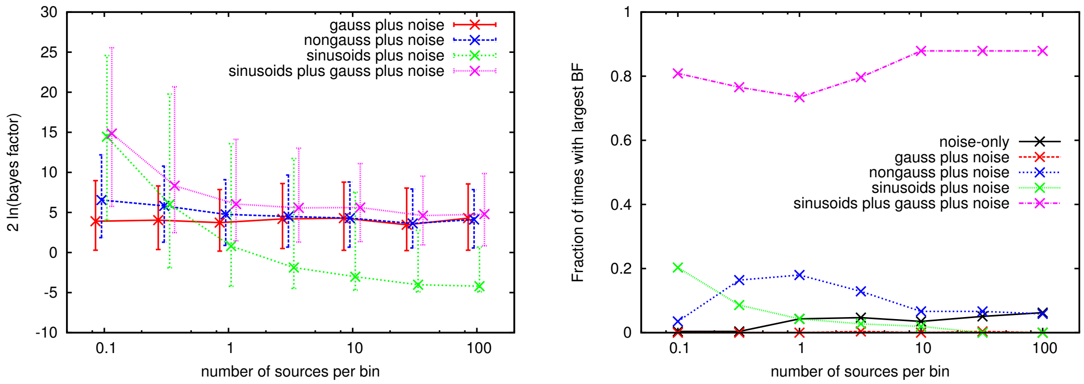


Figure 59: Left panel: Bayes factor 80% quantile intervals for the four different signal-plus-noise models relative to the noise-only model as a function of the number of sources per bin. Right panel: Fraction of time that the different models had the largest Bayes factor for the different simulations. (Figure taken from [44].)

are not essential for this approach; the calculation for separated and misalinged detectors with non-isotropic response functions can also be done [153].

Let's begin then by denoting the output of the four detectors $I = 1, 2, 3, 4$ in the Fourier domain by

$$\tilde{d}_I = \tilde{n}_I + \tilde{h}\,, \qquad \tilde{h} = \tilde{g} + \sum_{i=1}^{n} \tilde{b}_i\,, \tag{8.20}$$

where $\tilde{n}_I$ denotes the noise in detector I and $\tilde{h}$ denotes the total gravitational-wave contribution, which has a Gaussian-stochastic component $\tilde{g}$, and a non-Gaussian component formed from the superposition of short-duration burst signals $\tilde{b}_i$, $i = 1, 2, \cdots, n$. We assume that the noise in the detectors are uncorrelated with one another and with the gravitational-wave signals, and that the individual gravitational-wave signals are also uncorrelated amongst themselves. The (random) number of bursts present in a particular segment of data is determined by a Poisson distribution

$$P(n) = \frac{\lambda^n e^{-\lambda}}{n!}\,, \tag{8.21}$$

where

$$\lambda = \langle n \rangle = \sum_{n=0}^{\infty} n P(n)\,, \tag{8.22}$$

is the expected number of bursts in segment duration $T_{\rm seg}$. The 4th-order combination of data that we consider is

$$\mathcal{K} \equiv \langle \tilde{d}_1 \tilde{d}_2 \tilde{d}_3^* \tilde{d}_4^* \rangle - \langle \tilde{d}_1 \tilde{d}_2 \rangle \langle \tilde{d}_3^* \tilde{d}_4^* \rangle - \langle \tilde{d}_1 \tilde{d}_3^* \rangle \langle \tilde{d}_2 \tilde{d}_4^* \rangle - \langle \tilde{d}_1 \tilde{d}_4^* \rangle \langle \tilde{d}_2 \tilde{d}_3^* \rangle\,, \tag{8.23}$$

where angle brackets $\langle\ \rangle$ can be thought of as either expectation value (i.e., ensemble average) or as an average over the Fourier components of the data, i.e., as an *estimator* of the expected correlations. Since the noise in the detectors are uncorrelated with everything, the only contributions to $\mathcal{K}$ will come from expectation values of products of $\tilde{h} = \tilde{g} + \sum_i \tilde{b}_i$ with itself. Calculating the quadratic terms that enter (8.23), we find:

$$\begin{aligned} \langle \tilde{d}_I \tilde{d}_J \rangle &= \langle \tilde{g}\tilde{g} \rangle + \lambda \langle \tilde{b}\tilde{b} \rangle \,, \\ \langle \tilde{d}_I \tilde{d}_J^* \rangle &= \langle \tilde{g}\tilde{g}^* \rangle + \lambda \langle \tilde{b}\tilde{b}^* \rangle \,, \end{aligned} \tag{8.24}$$

where we used

$$\left\langle \sum_i \sum_j \tilde{b}_i \tilde{b}_j \right\rangle = \left\langle \sum_i \tilde{b}_i \tilde{b}_i \right\rangle = \lambda \langle \tilde{b}\tilde{b} \rangle \,, \tag{8.25}$$

which assumes that all the bursts have the same mean-square value, $\langle \tilde{b}_i \tilde{b}_i \rangle \equiv \langle \tilde{b}\tilde{b} \rangle$. For the 4th-order term, we find:

$$\begin{aligned} \langle \tilde{d}_1 \tilde{d}_2 \tilde{d}_3^* \tilde{d}_4^* \rangle = \langle \tilde{g}\tilde{g}\tilde{g}^*\tilde{g}^* \rangle &+ \lambda^2 \left[|\langle \tilde{b}\tilde{b} \rangle|^2 + 2\langle \tilde{b}\tilde{b}^* \rangle^2 \right] \\ &+ \lambda \left[\langle \tilde{b}\tilde{b}\tilde{b}^*\tilde{b}^* \rangle + \langle \tilde{g}\tilde{g} \rangle \langle \tilde{b}\tilde{b} \rangle^* + \langle \tilde{g}\tilde{g} \rangle^* \langle \tilde{b}\tilde{b} \rangle + 4\langle \tilde{g}\tilde{g}^* \rangle \langle \tilde{b}\tilde{b}^* \rangle \right] . \end{aligned} \tag{8.26}$$

Substituting these results back into expression (8.23) yields:

$$\mathcal{K} = \lambda \langle \tilde{b}\tilde{b}\tilde{b}^*\tilde{b}^* \rangle \,, \tag{8.27}$$

where we used

$$\langle \tilde{g}\tilde{g}\tilde{g}^*\tilde{g}^* \rangle - |\langle \tilde{g}\tilde{g} \rangle|^2 - 2\langle \tilde{g}\tilde{g}^* \rangle^2 = 0 \,, \tag{8.28}$$

for the Gaussian-stochastic signal component $\tilde{g}$. Thus, both the detector noise and the Gaussian-stochastic component of the signal have dropped out of the expression for $\mathcal{K}$, leaving only the contribution from the non-Gaussian component of the background.

As mentioned already, the above calculation can be extended to the case of separated and misaligned detectors [153]. In so doing, one obtains expressions for *generalized* (4th-order) overlap functions, which are sky-averages of the product of the response functions for four different detectors. The expected value of the 4th-order detection statistic for this more general analysis involves generalized overlap functions for both the (squared) overall intensity and circular polarization components of the non-Gaussian background. We will discuss circular polarization in the following section, but in the simpler context of Gaussian-stationary isotropic backgrounds. Readers should see [153] for more details regarding circular polarization in the context of non-Gaussian stochastic signals discussed above.

8.2 Circular polarization

Up until now, we have only considered *unpolarized* stochastic backgrounds. That is, we have assumed that the gravitational-wave power in the + and × polarization modes are equal (on average) and are statistically independent of one another (i.e., there are no correlations between the + and × polarization modes). It is possible, however, for some processes in the early universe to give rise to *parity violations* [21], which would manifest themselves as an asymmetry in the amount of right and left *circularly* polarized gravitational waves. Following [155, 156], we now describe how to generalize our cross-correlation methods to look for evidence of circular polarization in a stochastic background.

8.2.1 Polarization correlation matrix

Let's start by writing down the quadratic expectation values for the Fourier components $h_{ab}(f,\hat{n})$ of the metric perturbations $h_{ab}(t,\vec{x})$ for a *polarized anisotropic* Gaussian-stationary background. (We will restrict attention to isotropic backgrounds later on.) It turns out that these expectation values can be written in terms of the *Stoke's parameters* I, Q, U and V, which are defined for a monochromatic plane gravitational wave in Appendix A. If we expand $h_{ab}(f,\hat{n})$ in terms of the *linear* polarization basis tensors $e^A_{ab}(\hat{n})$, where $A=\{+,\times\}$, we have

$$\langle h_A(f,\hat{n})h^*_{A'}(f',\hat{n}')\rangle = \frac{1}{2}S_h^{AA'}(f,\hat{n})\delta(f-f')\delta^2(\hat{n},\hat{n}')\,, \tag{8.29}$$

where

$$S_h^{AA'}(f,\hat{n}) = \frac{1}{2}\begin{bmatrix} I(f,\hat{n})+Q(f,\hat{n}) & U(f,\hat{h})-iV(f,\hat{n}) \\ U(f,\hat{n})+iV(f,\hat{n}) & I(f,\hat{n})-Q(f,\hat{n})\end{bmatrix}. \tag{8.30}$$

If instead we expand $h_{ab}(f,\hat{n})$ in terms of the *circular* polarization basis tensors $e^C_{ab}(\hat{n})$, where $C=\{R,L\}$, then

$$\langle h_C(f,\hat{n})h^*_{C'}(f',\hat{n}')\rangle = \frac{1}{2}S_h^{CC'}(f,\hat{n})\delta(f-f')\delta^2(\hat{n},\hat{n}')\,, \tag{8.31}$$

where

$$S_h^{CC'}(f,\hat{n}) = \frac{1}{2}\begin{bmatrix} I(f,\hat{n})+V(f,\hat{n}) & Q(f,\hat{h})-iU(f,\hat{n}) \\ Q(f,\hat{n})+iU(f,\hat{n}) & I(f,\hat{n})-V(f,\hat{n})\end{bmatrix}. \tag{8.32}$$

This second representation of the polarization correlation matrix is sometimes more convenient when one is searching for evidence of circular polarization in the background, as V is a measure of a possible asymmetry between the right and left circular polarization components:

$$\langle h_R(f,\hat{n})h^*_{R'}(f',\hat{n}')\rangle - \langle h_L(f,\hat{n})h^*_{L'}(f',\hat{n}')\rangle = \frac{1}{2}V(f,\hat{n})\delta(f-f')\delta^2(\hat{n},\hat{n}')\,. \tag{8.33}$$

The factor of 1/2 on the right-hand side of the above equation, as compared to (A.16), is for one-sided power spectra.

As discussed in Appendix A, the Stokes' parameters I and V are ordinary scalar (spin 0) fields on the sphere, while Q and U transform like spin 4 fields under a rotation of the unit vectors $\{\hat{l},\hat{m}\}$ tangent to the sphere. Thus, I and V can be written as linear combinations of the ordinary spherical harmonics $Y_{lm}(\hat{n})$:

$$\begin{aligned} I(f,\hat{n}) &= \sum_{l=0}^{\infty}\sum_{m=-l}^{l} I_{lm}(f)Y_{lm}(\hat{n})\,, \\ V(f,\hat{n}) &= \sum_{l=0}^{\infty}\sum_{m=-l}^{l} V_{lm}(f)Y_{lm}(\hat{n})\,, \end{aligned} \tag{8.34}$$

while $Q \pm iU$ can be written as linear combination of the spin-weighted ± 4 spherical harmonics ${}_{\pm 4}Y_{lm}(\hat{n})$:

$$Q(f,\hat{n}) \pm iU(f,\hat{n}) = \sum_{l=4}^{\infty} \sum_{m=-l}^{l} C_{lm}^{\pm}(f)\,{}_{\pm 4}Y_{lm}(\hat{n})\,. \tag{8.35}$$

Note that the expansions for $Q \pm iU$ start at $l = 4$, which means that the Q, U components of the polarization correlation matrix vanish if the background is isotropic (i.e., has only a contribution from the monopole $l = 0$, $m = 0$). So for simplicity, we will restrict our attention to polarized *isotropic backgrounds*, for which the circular polarization correlation matrix becomes diagonal and the quadratic expectation values reduce to:

$$\langle h_C(f,\hat{n}) h_{C'}^*(f',\hat{n}')\rangle = \frac{1}{8\pi} S_h^C(f)\delta_{CC'}\delta(f-f')\delta^2(\hat{n},\hat{n}')\,, \tag{8.36}$$

where

$$\begin{aligned} S_h^R(f) &\equiv \frac{1}{2}(I(f)+V(f))\,,\\ S_h^L(f) &\equiv \frac{1}{2}(I(f)-V(f))\,. \end{aligned} \tag{8.37}$$

Note that

$$S_h^R(f) + S_h^L(f) = I(f) \equiv S_h(f)\,, \tag{8.38}$$

which is just the total strain power spectral density for the gravitational-wave background.

8.2.2 Overlap functions

Given (8.36), we are now in a position to calculate the expected value of the product of the Fourier transforms of the response of two detectors I and J to such a background. Similar to (5.9), we can write the response of detector I as

$$\tilde{h}_I(f) = \int d^2\Omega_{\hat{n}}\ \left(R^R(f,\hat{n})h_R(f,\hat{n}) + R^L(f,\hat{n})h_L(f,\hat{n})\right)\,, \tag{8.39}$$

where R, L label the right and left circular polarization states for both the Fourier components and the detector response functions. Writing down a similar expression for the response of detector J, and using (8.36) to evaluate the expected value of the product of the responses, we find

$$\langle \tilde{h}_I(f)\tilde{h}_J^*(f')\rangle = \frac{1}{2}\delta(f-f')\left[\Gamma_{IJ}^{(I)}(f)I(f) + \Gamma_{IJ}^{(V)}(f)V(f)\right]\,, \tag{8.40}$$

where

$$\begin{aligned} \Gamma_{IJ}^{(I)}(f) &\equiv \frac{1}{8\pi}\int d^2\Omega_{\hat{n}}\ \left[R_I^R(f,\hat{n})R_J^{R*}(f,\hat{n}) + R_I^L(f,\hat{n})R_J^{L*}(f,\hat{n})\right]\,,\\ \Gamma_{IJ}^{(V)}(f) &\equiv \frac{1}{8\pi}\int d^2\Omega_{\hat{n}}\ \left[R_I^R(f,\hat{n})R_J^{R*}(f,\hat{n}) - R_I^L(f,\hat{n})R_J^{L*}(f,\hat{n})\right]\,, \end{aligned} \tag{8.41}$$

are the overlap functions for the I and V Stokes parameters for a polarized isotropic stochastic background. Using

$$\begin{aligned} R^R &= \frac{1}{\sqrt{2}}\left(R^+ + iR^\times\right) \,, \\ R^L &= \frac{1}{\sqrt{2}}\left(R^+ - iR^\times\right) \,, \end{aligned} \tag{8.42}$$

we can also write the above overlap functions as

$$\begin{aligned} \Gamma^{(I)}_{IJ}(f) &\equiv \frac{1}{8\pi}\int d^2\Omega_{\hat{n}} \,\left[R^+_I(f,\hat{n})R^{+*}_J(f,\hat{n}) + R^\times_I(f,\hat{n})R^{\times *}_J(f,\hat{n})\right] \,, \\ \Gamma^{(V)}_{IJ}(f) &\equiv \frac{i}{8\pi}\int d^2\Omega_{\hat{n}} \,\left[R^+_I(f,\hat{n})R^{+*}_J(f,\hat{n}) - R^\times_I(f,\hat{n})R^{\times *}_J(f,\hat{n})\right] \,. \end{aligned} \tag{8.43}$$

Note that $\Gamma^{(I)}_{IJ}(f)$ is identical to the ordinary overlap function $\Gamma_{IJ}(f)$ for an isotropic background (5.39).

Figure 60 show plots of the I and V overlap functions for the LIGO-Virgo detector pairs, using the small-antenna limit for the strain response functions. The overlap functions have been normalized (5.43) so that $\gamma^{(I)}_{IJ}(f) = 1$ for colocated and coaligned detectors. Similar plots can be made for other interferometer pairs, by simply using the appropriate response functions for those detectors.

NOTE: For pulsar timing, one can show that $\Gamma^{(V)}_{IJ}(f) = 0$ for any pair of pulsars. This means that one cannot detect the presence of a circularly polarized stochastic background using a pulsar timing array if one restricts attention to just the isotropic component of the background. One must include higher-order multipoles in the analysis—i.e., do an *anisotropic* search as discussed in Section 7. Such an analysis for anisotropic polarized backgrounds using pulsar timing arrays is given in [101]. In that paper, they extend the analysis of [120] to include circular polarization. Please see [101] for additional details.

8.2.3 Component separation: ML estimates of I and V

As shown in [155, 156], in order to separate the $I(f)$ and $V(f)$ contributions to a polarized isotropic background, we will need to analyze data from at least two independent baselines (so three or more detectors). In what follows, we will use the notation $\alpha = 1, 2, \cdots, N_b$ to denote the individual baselines (detector pairs) and α_1, α_2 to denote the two detectors that constitute that baseline. The formalism we adopt here is similar to that for constructing maximum-likelihood estimators of gravitational-wave power for unpolarized anisotropic backgrounds (Section 7.3). For a general discussion of component separation for isotropic backgrounds, see [132].

As usual, we begin by cross-correlating the data from pairs of detectors for the independent baselines:

$$\hat{C}_\alpha(f) \equiv \frac{2}{T}\,\tilde{d}_{\alpha_1}(f)\tilde{d}^*_{\alpha_2}(f)\,, \tag{8.44}$$

where

$$\tilde{d}_{\alpha_I}(f) = \tilde{h}_{\alpha_I}(\tilde{f}) + \tilde{n}_{\alpha_I}(f)\,, \quad I = 1, 2\,, \tag{8.45}$$

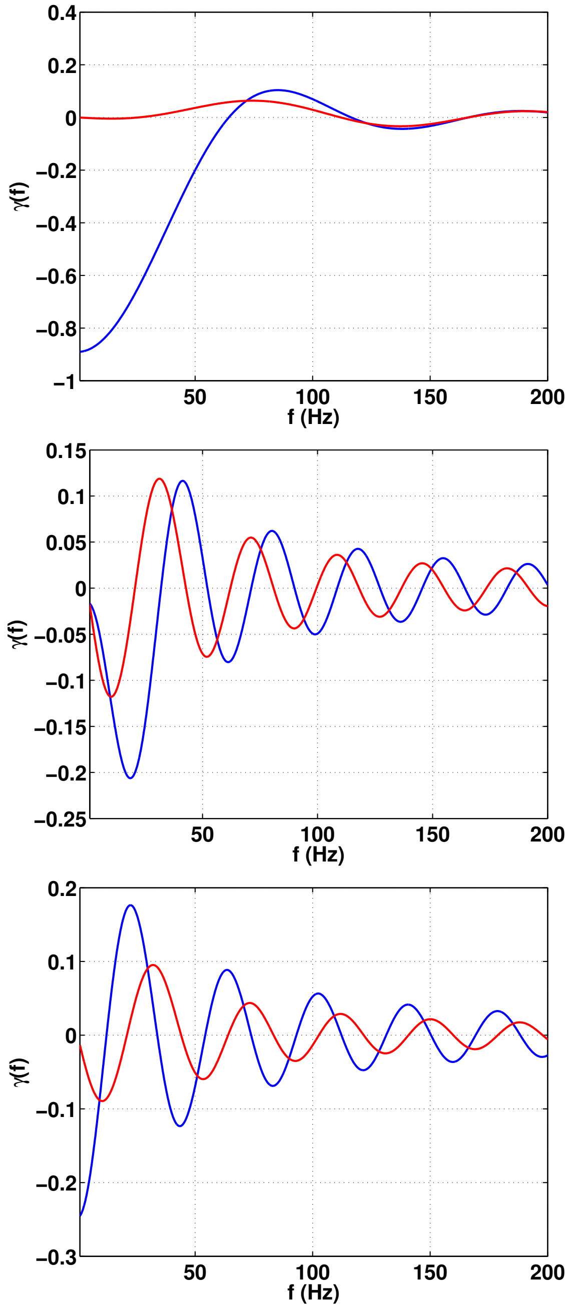


Figure 60: Normalized overlap functions for the I and V Stokes' parameters for the LIGO Hanford-LIGO Livingston detector pair (top panel); for the LIGO Hanford-Virgo detector pair (middle panel); for the LIGO Livingston-Virgo detector pair (bottom panel). The I overlap functions are shown in blue; the V overlap functions are shown in red. Note the change in scale of the vertical axes.

are the Fourier transforms of the time-domain data $d_{\alpha_I}(t)$, and T is the duration of the data. Assuming that the noise in the individual detectors are uncorrelated with one another, we can easily calculate the expected value of $\hat{C}_\alpha(f)$ using our previous result (8.40). The result is

$$\langle \hat{C}_\alpha(f)\rangle = \Gamma_\alpha^{(I)}(f)I(f) + \Gamma_\alpha^{(V)}(f)V(f)\,. \tag{8.46}$$

We will write this equation abstractly as a matrix equation

$$\langle \hat{C}\rangle = M\mathcal{S}\,, \tag{8.47}$$

where

$$\hat{C} = \begin{bmatrix} \hat{C}_1 \\ \hat{C}_2 \\ \vdots \\ \hat{C}_{N_b} \end{bmatrix}, \quad M \equiv \begin{bmatrix} \Gamma_1^{(I)} & \Gamma_1^{(V)} \\ \Gamma_2^{(I)} & \Gamma_2^{(V)} \\ \vdots & \vdots \\ \Gamma_{N_b}^{(I)} & \Gamma_{N_b}^{(V)} \end{bmatrix}, \quad \mathcal{S} \equiv \begin{bmatrix} I \\ V \end{bmatrix}. \tag{8.48}$$

In this notation, $\hat{C}$ is an $N_f N_b \times 1$ data vector, M is an $N_f N_b \times 2N_f$ detector network response matrix, and $\mathcal{S}$ is an $2N_f \times 1$ vector containing the unknown Stokes parameters, which we want to estimate from the data.[20]

We also need an expression for the noise covariance matrix $\mathcal{N}$ for the cross-correlated data $\hat{C}$. In the weak-signal limit, the covariance matrix is *approximately diagonal* with matrix elements

$$\begin{aligned} \mathcal{N}_{\alpha\alpha'}(f,f') &\equiv \langle \hat{C}_\alpha(f)\hat{C}^*_{\alpha'}(f')\rangle - \langle \hat{C}_\alpha(f)\rangle\langle \hat{C}^*_{\alpha'}(f')\rangle \\ &\approx \delta_{\alpha\alpha'}\delta_{ff'}P_{n_{\alpha_1}}(f)P_{n_{\alpha_2}}(f)\,, \end{aligned} \tag{8.49}$$

where $P_{n_{\alpha_I}}(f)$ are the one-sided power spectral densities of the noise in the detectors. If we treat the noise power spectra as known quantities (or if we estimate the noise power spectra from the auto-correlated output of each detector), we can write down a likelihood function for the cross-correlated data given the signal model (8.47). Assuming a Gaussian-stationary distribution for the noise, we have

$$p(\hat{C}|\mathcal{S}) \propto \exp\left[-\frac{1}{2}(\hat{C}-M\mathcal{S})^\dagger \mathcal{N}^{-1}(\hat{C}-M\mathcal{S})\right]. \tag{8.50}$$

This likelihood has exactly the same form as that in (7.32), so the maximum-likehood estimators for the Stokes' parameters $\mathcal{S} = [I,V]^T$ also have the same form:

$$\hat{\mathcal{S}} = F^{-1}X\,, \tag{8.51}$$

where

$$F \equiv M^\dagger \mathcal{N}^{-1} M\,, \qquad X \equiv M^\dagger \mathcal{N}^{-1}\hat{C}\,, \tag{8.52}$$

[20] At times it will be convenient to think of M as an $N_f \times N_f$ block diagonal matrix with $N_b \times 2$ blocks, one for each frequency. At other times, it will be convenient to think of M as an $N_b \times 2$ block matrix with diagonal $N_f \times N_f$ blocks. The calculations we need to do usually determine which representation is most appropriate. (Similar statements can be made for the vectors $\hat{C}$ and $\mathcal{S}$.)

with M and $\mathcal{N}$ given above. As before, inverting F may require some sort of regularization, e.g., using singular-value decomposition (Section 7.3.5). If that's the case then F^{-1} should be replaced in the above formula by its pseudo-inverse F^{+}. The uncertainty in the maximum likelihood recoved values is given by the covariance matrix

$$\langle \hat{\mathcal{S}}\hat{\mathcal{S}}^{\dagger}\rangle - \langle \hat{\mathcal{S}}\rangle\langle \hat{\mathcal{S}}^{\dagger}\rangle \approx F^{-1}\,, \tag{8.53}$$

where we are again assuming the weak-signal limit.

8.2.4 Example: Component separation for two baselines

As an explicit example, we now write down the maximum-likelihood estimators for the Stokes' parameters $\mathcal{S} = [I, V]^T$ for a detector network consisting of two baselines α and β. For this case, the detector network response matrix M is a square $2N_f \times 2N_f$ matrix, which we assume has non-zero determinant. Then it follows simply from the definitions (8.52) of F and X that

$$\hat{\mathcal{S}} = F^{-1}X = M^{-1}\hat{C}\,, \tag{8.54}$$

for which

$$\begin{aligned}
\hat{I}(f) &= \left(\Gamma_\alpha^{(I)}\Gamma_\beta^{(V)} - \Gamma_\beta^{(I)}\Gamma_\alpha^{(V)}\right)^{-1}\left[\Gamma_\beta^{(V)}\hat{C}_\alpha - \Gamma_\alpha^{(V)}\hat{C}_\beta\right],\\
\hat{V}(f) &= \left(\Gamma_\alpha^{(I)}\Gamma_\beta^{(V)} - \Gamma_\beta^{(I)}\Gamma_\alpha^{(V)}\right)^{-1}\left[-\Gamma_\beta^{(I)}\hat{C}_\alpha + \Gamma_\alpha^{(I)}\hat{C}_\beta\right].
\end{aligned} \tag{8.55}$$

The marginalized uncertainties in these estimates are obtained by taking the diagonal elements of the inverse of the Fisher matrix:

$$\begin{aligned}
\sigma^2_{\hat{I}} = (F^{-1})_{II} &= \frac{N_\alpha\,(\Gamma_\beta^{(V)})^2 + N_\beta\,(\Gamma_\alpha^{(V)})^2}{\left(\Gamma_\alpha^{(I)}\Gamma_\beta^{(V)} - \Gamma_\beta^{(I)}\Gamma_\alpha^{(V)}\right)^2}\,,\\
\sigma^2_{\hat{V}} = (F^{-1})_{VV} &= \frac{N_\alpha\,(\Gamma_\beta^{(I)})^2 + N_\beta\,(\Gamma_\alpha^{(I)})^2}{\left(\Gamma_\alpha^{(I)}\Gamma_\beta^{(V)} - \Gamma_\beta^{(I)}\Gamma_\alpha^{(V)}\right)^2}\,,
\end{aligned} \tag{8.56}$$

where N_α, N_β defined by $N_\alpha(f) \equiv P_{n_{\alpha_1}}(f)P_{n_{\alpha_2}}(f)$ (and similarly for N_β) is a diagonal element of the noise covariance matrix $\mathcal{N}$ (8.49).

8.2.5 Effective overlap functions for I and V for multiple baselines

The above expressions for the uncertainties in the estimates of I and V can easily be extended to the case of an arbitray number of baselines $\alpha = 1, 2, \cdots, N_b$. For multiple baselines with noise spectra $N_\alpha(f) \equiv P_{n_{\alpha_1}}(f)P_{n_{\alpha_2}}(f)$, one can show that

$$F = \begin{bmatrix} \sum_\alpha N_\alpha^{-1}(\Gamma_\alpha^{(I)})^2 & \sum_\alpha N_\alpha^{-1}\Gamma_\alpha^{(I)}\Gamma_\alpha^{(V)} \\ \sum_\alpha N_\alpha^{-1}\Gamma_\alpha^{(V)}\Gamma_\alpha^{(I)} & \sum_\alpha N_\alpha^{-1}(\Gamma_\alpha^{(V)})^2 \end{bmatrix}. \tag{8.57}$$

Let us assume that the determinant of the 2×2 matrices for each frequency (which we will denote by $\bar{F}$) are not equal to zero. Then

$$\begin{aligned}\sigma^2_{\hat{I}} &= (\bar{F}^{-1})_{II} = \frac{1}{\det(\bar{F})} \sum_\alpha N_\alpha^{-1} (\Gamma_\alpha^{(V)})^2 \,, \\ \sigma^2_{\hat{V}} &= (\bar{F}^{-1})_{VV} = \frac{1}{\det(\bar{F})} \sum_\alpha N_\alpha^{-1} (\Gamma_\alpha^{(I)})^2 \,.\end{aligned} \tag{8.58}$$

Following [156], we can now define *effective* overlap functions for I and V associated with a multibaseline detector network by basically inverting the above uncertainties. For simplicity, we will assume that the noise power spectra for the detectors are equal to one another so that $N_\alpha \equiv N$ can be factored out of the above expressions. We then define

$$\begin{aligned}\Gamma^{(I)}_{\text{eff}}(f) &\equiv \sqrt{N}\sigma^{-1}_{\hat{I}} = \left(\frac{N^2 \det(\bar{F})}{\sum_\alpha (\Gamma_\alpha^{(V)})^2}\right)^{1/2} \,, \\ \Gamma^{(V)}_{\text{eff}}(f) &\equiv \sqrt{N}\sigma^{-1}_{\hat{V}} = \left(\frac{N^2 \det(\bar{F})}{\sum_\alpha (\Gamma_\alpha^{(I)})^2}\right)^{1/2} \,.\end{aligned} \tag{8.59}$$

These quantities give us an indication of how sensitive the multibaseline network is to extracting the I and V components of the background. Plots of $\Gamma^{(I)}_{\text{eff}}(f)$ and $\Gamma^{(V)}_{\text{eff}}(f)$ are shown in Figure 61 for the multibaseline network formed from the LIGO Hanford, LIGO Livingston, and Virgo detectors. Recall that the overlap functions for the individual

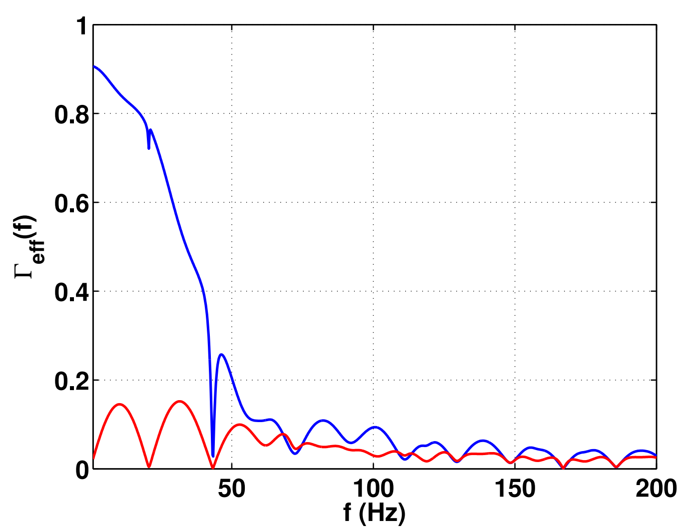


Figure 61: Effective overlap functions for I and V for the multibaseline network formed from the LIGO Hanford, LIGO Livingston, and Virgo detectors. $\Gamma^{(I)}_{\text{eff}}(f)$ is shown in blue; $\Gamma^{(V)}_{\text{eff}}(f)$ is shown in red.

detectors pairs are shown in Figure 60. Dips in sensitivity correspond to frequencies where the determinant of $\bar{F}$ is zero (or close to zero).

8.3 non-GR polarization modes: Preliminaries

In a general metric theory of gravity, there are six possible polarization modes: The standard $+$ and $\times$ *tensor* modes predicted by general relativity (GR); two *vector* (or "shear") modes, which we will denote by X and Y; and two *scalar* modes: a "breathing" mode B and a pure longitudinal mode L (see, e.g., [130]). The tensor and breathing modes are *transverse* to the direction of propagation, while the two vector modes and the scalar longitudinal mode have *longitudinal* components (parallel to the direction of propagation). See Figure 62.

In terms of the orthonormal vectors $\{\hat{n}, \hat{l}, \hat{m}\}$ defined by (2.4), the polarization basis tensors for the six different polarization modes are:

$$\begin{aligned} e^{+}_{ab}(\hat{n}) &= \hat{l}_a\hat{l}_b - \hat{m}_a\hat{m}_b\,, & e^{\times}_{ab}(\hat{n}) &= \hat{l}_a\hat{m}_b + \hat{m}_a\hat{l}_b\,. \\ e^{X}_{ab}(\hat{n}) &= \hat{l}_a\hat{n}_b + \hat{n}_a\hat{l}_b\,, & e^{Y}_{ab}(\hat{n}) &= \hat{m}_a\hat{n}_b + \hat{n}_a\hat{m}_b\,, \\ e^{B}_{ab}(\hat{n}) &= \hat{l}_a\hat{l}_b + \hat{m}_a\hat{m}_b\,, & e^{L}_{ab}(\hat{n}) &= \sqrt{2}\,\hat{n}_a\hat{n}_b\,. \end{aligned} \tag{8.60}$$

We will denote these collectively as $e^{A}_{ab}(\hat{n})$, where $A = \{+, \times, X, Y, B, L\}$. In a coordinate system where $\hat{n}$ points along the z-axis, and $\hat{l}$ and $\hat{m}$ point along the x and y axes, the polarization tensors can be represented by the following 3×3 matrices:

$$\begin{aligned} e^{+}_{ab} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, & e^{\times}_{ab} &= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \\ e^{X}_{ab} &= \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, & e^{Y}_{ab} &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \\ e^{B}_{ab} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, & e^{L}_{ab} &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix}. \end{aligned} \tag{8.61}$$

8.3.1 Tranformation of the polarization tensors under a rotation about $\hat{n}$

We have already seen (Appendix A) that under a rotation of the unit vectors $\{\hat{l}, \hat{m}\}$ by an angle ψ around $\hat{n}$, the polarization tensors $e^{+}_{ab}(\hat{n})$, $e^{\times}_{ab}(\hat{n})$ transform to:

$$\begin{aligned} \epsilon^{+}_{ab}(\hat{n}, \psi) &= \cos 2\psi\, e^{+}_{ab}(\hat{n}) + \sin 2\psi\, e^{\times}_{ab}(\hat{n})\,, \\ \epsilon^{\times}_{ab}(\hat{n}, \psi) &= -\sin 2\psi\, e^{+}_{ab}(\hat{n}) + \cos 2\psi\, e^{\times}_{ab}(\hat{n})\,. \end{aligned} \tag{8.62}$$

This reflects the spin 2 nature of the tensor modes $+$, $\times$ in general relativity. Similarly, under the same rotation, the polarization tensors $e^{X}_{ab}(\hat{n})$, $e^{Y}_{ab}(\hat{n})$ transform to:

$$\begin{aligned} \epsilon^{X}_{ab}(\hat{n}, \psi) &= \cos \psi\, e^{X}_{ab}(\hat{n}) + \sin \psi\, e^{Y}_{ab}(\hat{n})\,, \\ \epsilon^{Y}_{ab}(\hat{n}, \psi) &= -\sin \psi\, e^{X}_{ab}(\hat{n}) + \cos \psi\, e^{Y}_{ab}(\hat{n})\,, \end{aligned} \tag{8.63}$$

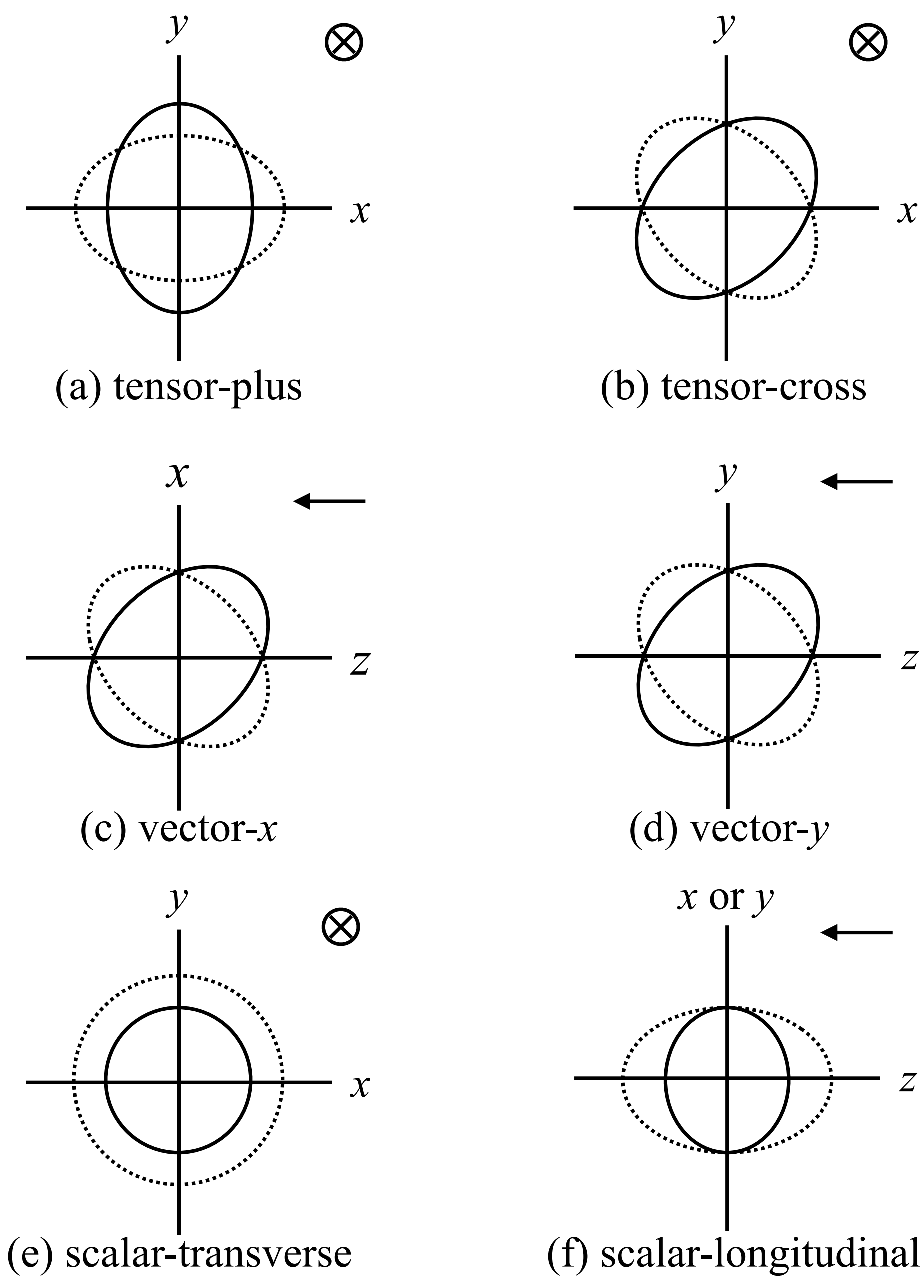


Figure 62: Graphical representation of the six different polarization modes. The circle with a cross or arrow represents the direction of propagation of the gravitational wave. The solid and dotted circles and ellipse denote deformations to a ring of particles 180° out of phase with one another. Adapted from Figure 1 in [130].

while $e^B_{ab}(\hat n)$, $e^L_{ab}(\hat n)$ are left unchanged:

$$\begin{aligned}\epsilon^B_{ab}(\hat n,\psi) &= e^B_{ab}(\hat n)\,,\\ \epsilon^L_{ab}(\hat n,\psi) &= e^L_{ab}(\hat n)\,.\end{aligned} \tag{8.64}$$

These last two transformations correspond to the spin 1 nature of the vector modes X, Y, and the spin 0 nature of the scalar modes B, L.

8.3.2 Polarization and spherical harmonic basis expansions

For the tensor modes $+$, $\times$, we found (Section 2.2.2) that it was convenient to expand the Fourier components $h_{ab}(f,\hat k)$ of the metric perturbations $h_{ab}(t,\vec x)$ in terms of either the polarization basis tensors:

$$h_{ab}(f,\hat n) = h_+(f,\hat n)e^+_{ab}(\hat n) + h_\times(f,\hat n)e^\times_{ab}(\hat n)\,, \tag{8.65}$$

or the gradient and curl tensor spherical harmonics:

$$h_{ab}(f,\hat n) = \sum_{l=2}^{\infty}\sum_{m=-l}^{l}\left[a^G_{(lm)}(f)Y^G_{(lm)ab}(\hat n) + a^C_{(lm)}(f)Y^C_{(lm)ab}(\hat n)\right]\,. \tag{8.66}$$

Recall that Y^G and Y^C are related to spin-weight ± 2 spherical harmonics as described in Appendices F and D. For the vector and scalar modes we can perform similar expansions, provided we use appropriately defined tensor spherical harmonics, which have the proper transformation property under rotations. For the vector modes X, Y, we need to use the *vector-gradient* and *vector-curl* spherical harmonics Y^{V_G}, Y^{V_C}, which are defined in terms of spin-weight ± 1 spherical harmonics (Appendices E and D). For the scalar modes, we can use

$$Y^B_{(lm)ab}(\hat n) \equiv \frac{1}{\sqrt 2}Y_{lm}(\hat n)e^B_{ab}(\hat n)\,,\qquad Y^L_{(lm)ab}(\hat n) \equiv \frac{1}{\sqrt 2}Y_{lm}(\hat n)e^L_{ab}(\hat n)\,, \tag{8.67}$$

which are defined in terms of ordinary (scalar) spherical harmonics. In terms of these definitions, we can write the expansions in compact form

$$h_{ab}(f,\hat n) = \sum_A h_A(f,\hat n)e^A_{ab}(\hat n)\,, \tag{8.68}$$

or

$$h_{ab}(f,\hat n) = \sum_P\sum_{(lm)} a^P_{(lm)}(f)Y^P_{(lm)ab}(\hat n)\,, \tag{8.69}$$

where $A = \{+,\times,X,Y,B,L\}$ and $P = \{G,C,V_G,V_C,B,L\}$ or some subsets thereof. Note that $\sum_{(lm)}$ is shorthand for

$$\sum_{l=2}^{\infty}\sum_{m=-l}^{l}\,,\qquad \sum_{l=1}^{\infty}\sum_{m=-l}^{l}\,,\qquad \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\,, \tag{8.70}$$

for the tensor, vector, and scalar modes, respectively.

8.3.3 Detector response

The detector response functions corresponding to the above two expansions (8.68) and (8.69) are:

$$R^A(f,\hat{n}) = R^{ab}(f,\hat{n})e^A_{ab}(\hat{n})\,, \tag{8.71}$$

and

$$R^P_{(lm)}(f) = \int d^2\Omega_{\hat{n}}\, R^{ab}(f,\hat{n})Y^P_{(lm)ab}(\hat{n})\,. \tag{8.72}$$

In terms of these response functions, the detector response (in the frequency domain) to a gravitational-wave background (2.1) is:

$$\tilde{h}(f) = \int d^2\Omega_{\hat{n}} \sum_A R^A(f,\hat{n})h_A(f,\hat{n})\,, \tag{8.73}$$

or

$$\tilde{h}(f) = \sum_P \sum_{(lm)} R^P_{(lm)}(f)a^P_{(lm)}(f)\,. \tag{8.74}$$

8.3.4 Searches for non-GR polarizations using different detectors

In the following sections we describe search methods for non-GR polarization modes using ground-based interferometers (Section 8.4) and pulsar timing arrays (Section 8.5). We will calculate antenna patterns, overlap functions, and discuss component separation for the tensor, vector, and scalar polarization modes. For ground-based interferometers, our discussion will be based on [130]. For pulsar timing arrays, see [109, 41, 71].

8.4 Searches for non-GR polarizations using ground-based detectors

We now describe cross-correlation searches for non-GR polarization modes using a network of ground-based laser interferometers. For additional details, see [130].

8.4.1 Response functions

For ground-based interferometers in the small antenna limit, the strain response functions $R^A(f,\hat{n})$ for the different polarization modes $A = \{+,\times,X,Y,B,L\}$ are given by

$$R^A(f,\hat{n}) \simeq \frac{1}{2}(u^a u^b - v^a v^b)e^A_{ab}(\hat{n})\,, \tag{8.75}$$

where $\hat{u}$, $\hat{v}$ are unit vectors pointing in the direction of the arms of the interferometer, and where we have chosen the origin of coordinates to be at the vertex of the interferometer. Note that there is no frequency dependence of the response function in the small-antenna limit. Assuming a 90° opening angle between the interferometer arms, and choosing a coordinate system such that $\hat{u}$ and $\hat{v}$ point in the $\hat{x}$ and $\hat{y}$ direction, we find

$$\begin{aligned}
R^+(\hat{n}) &= \frac{1}{2}(1+\cos^2\theta)\cos 2\phi\,, & R^\times(\hat{n}) &= -\cos\theta\sin 2\phi\,,\\
R^X(\hat{n}) &= \sin\theta\cos\theta\cos 2\phi\,, & R^Y(\hat{n}) &= -\sin\theta\sin 2\phi\,,\\
R^B(\hat{n}) &= -\frac{1}{2}\sin^2\theta\cos 2\phi\,, & R^L(\hat{n}) &= \frac{1}{\sqrt{2}}\sin^2\theta\cos 2\phi\,,
\end{aligned} \tag{8.76}$$

where we used (2.4) for our definition of $\{\hat{n}, \hat{l}, \hat{m}\}$.

Note that the response functions for the breathing and longitudinal modes differ only by a constant multiplicative factor of $-\sqrt{2}$. This degeneracy means that one will not be able to distinguish these two polarization modes using ground-based interferometers. Plots of the antenna patterns $|R^A(\hat{n})|$ for the six different polarization modes are shown in Figure 63. Note that the overall magnitude of the response gets smaller as one moves from tensor, to vector, to scalar polarization modes. In Figure 64, we plot the "peanut" antenna patterns for the response to unpolarized gravitational waves for the tensor, vector, and scalar modes, respectively. By unpolarized we simply mean that the incident gravitational waves have equal power in the $+$ and $\times$ polarizations for the tensor modes; equal power in the X and Y polarizations for the vector modes, and equal power in the B and L polarizations for the scalar modes.

8.4.2 Overlap functions

Similar to what we did in Section 8.2.2, let us assume that the stochastic background is *independently polarized*, but is otherwise Gaussian-stationary and isotropic. This means that the quadratic expectation values of the Fourier components of the metric perturbations can be written as

$$\langle h_A(f, \hat{n}) h^*_{A'}(f', \hat{n}') \rangle = \frac{1}{8\pi} S_h^A(f) \delta_{AA'} \delta(f - f') \delta^2(\hat{n}, \hat{n}') \,, \tag{8.77}$$

where $A = \{+, \times, X, Y, B, L\}$. The functions $S_h^A(f)$ are such that

$$\begin{aligned} S_h^{(T)}(f) &= S_h^+(f) + S_h^\times(f) \,, \\ S_h^{(V)}(f) &= S_h^X(f) + S_h^Y(f) \,, \\ S_h^{(S)}(f) &= S_h^B(f) + S_h^L(f) \,, \end{aligned} \tag{8.78}$$

are the one-sided strain spectral densities for the tensor, vector, and scalar modes individually. For simplicity, we will also assume that the tensor, vector, and scalar modes are individually unpolarized so that $S_h^+(f) = S_h^\times(f)$, $S_h^X(f) = S_h^Y(f)$, etc. All of these assumptions together define the stochastic signal model for this example.

The above expectation values (8.77) can now be used to calculate the expected value of the correlated response of two detectors to such a background. Writing the response of detector I as

$$\tilde{h}_I(f) = \int d^2\Omega_{\hat{n}} \sum_A R_I^A(f, \hat{n}) h_A(f, \hat{n}) \,, \tag{8.79}$$

it follows (as we have done many times before) that

$$\langle \tilde{h}_I(f) \tilde{h}^*_J(f') \rangle = \frac{1}{2} \delta(f - f') \left[\Gamma_{IJ}^{(T)}(f) S_h^{(T)}(f) + \Gamma_{IJ}^{(V)}(f) S_h^{(V)}(f) + \Gamma_{IJ}^{(S)}(f) S_h^{(S)}(f) \right] \,, \tag{8.80}$$

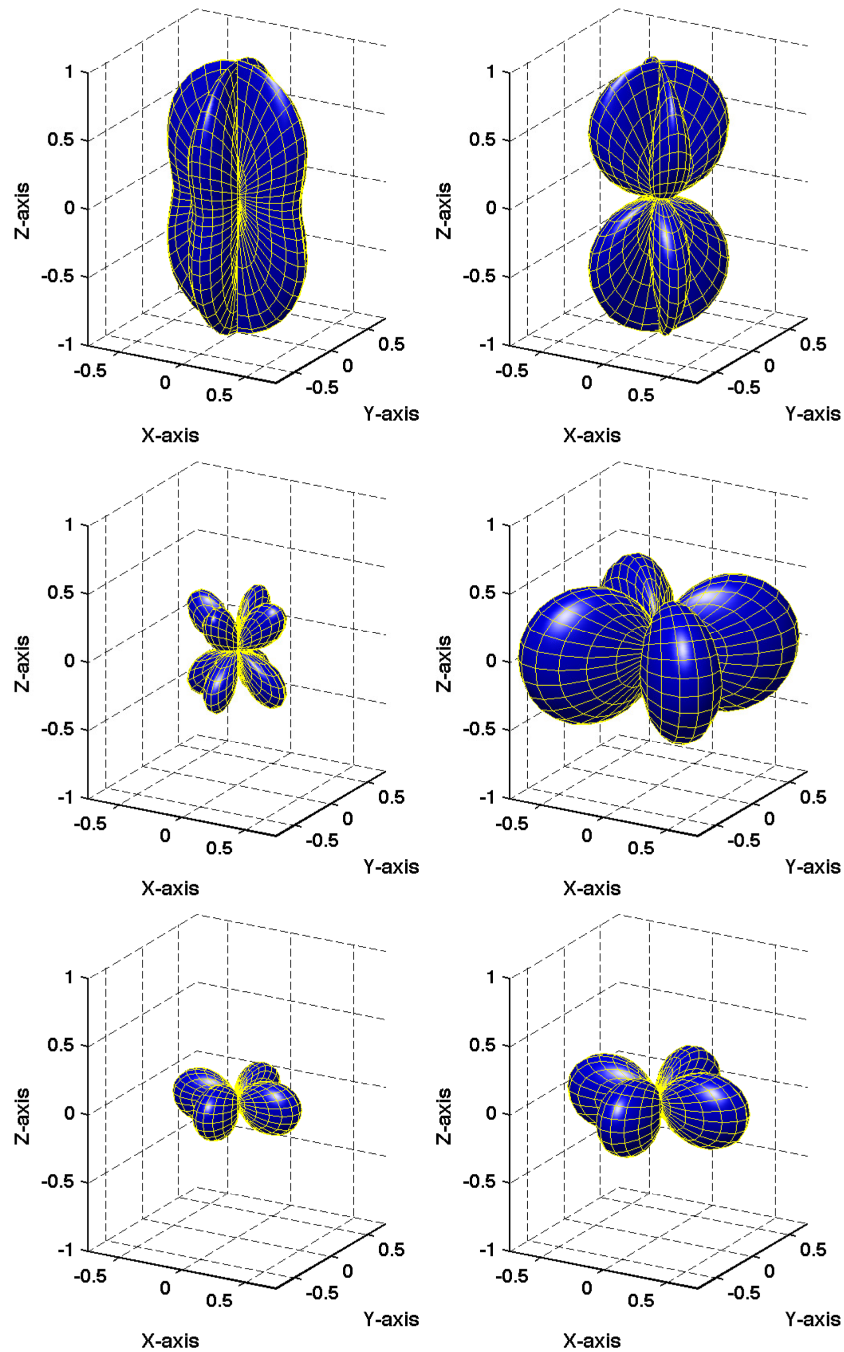


Figure 63: Antenna patterns for Michelson interferometer strain response $|R^A(\hat{n})|$ evaluated in small-antenna limit, $f = 0$. The top two plots correspond to the two tensor modes, $A = +, \times$. The middle two plots correspond to the two vector modes, $A = X, Y$. The bottom two plots correspond to the two scalar modes, $A = B, L$. The interferometer arms point in the $\hat{x}$ and $\hat{y}$ directions.

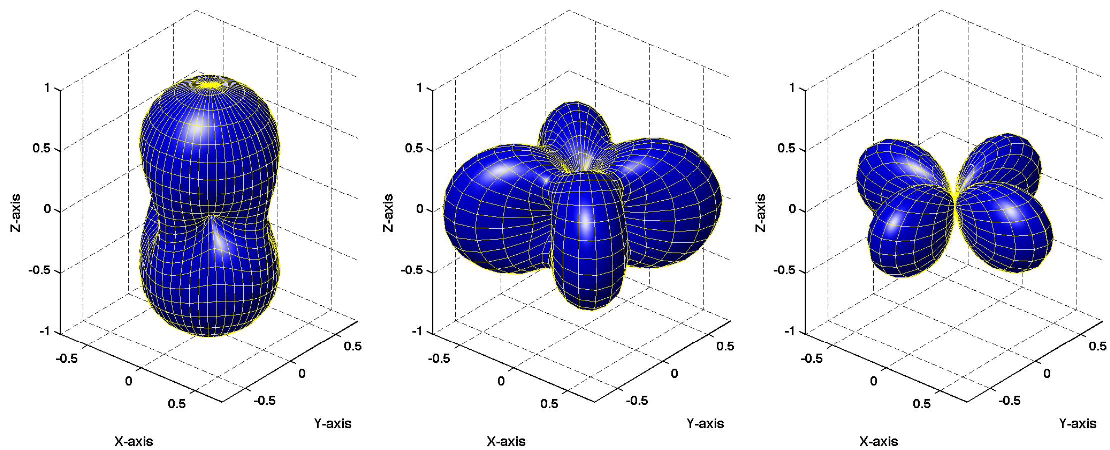


Figure 64: Antenna patterns for Michelson interferometer strain response to unpolarized gravitational waves for tensor (left plot), vector (middle plot), ans scalar modes (right plot), evaluated in the small antenna limit, $f = 0$. The interferometer arms point in the $\hat{x}$ and $\hat{y}$ directions.

where

$$\begin{aligned}
\Gamma^{(T)}_{IJ}(f) &\equiv \frac{1}{8\pi}\int d^2\Omega_{\hat{n}}\ \left[R^{+}_I(f,\hat{n})R^{+*}_J(f,\hat{n}) + R^{\times}_I(f,\hat{n})R^{\times *}_J(f,\hat{n})\right]\,,\\
\Gamma^{(V)}_{IJ}(f) &\equiv \frac{1}{8\pi}\int d^2\Omega_{\hat{n}}\ \left[R^{X}_I(f,\hat{n})R^{X*}_J(f,\hat{n}) + R^{Y}_I(f,\hat{n})R^{Y*}_J(f,\hat{n})\right]\,,\\
\Gamma^{(S)}_{IJ}(f) &\equiv \frac{1}{8\pi}\int d^2\Omega_{\hat{n}}\ \left[R^{B}_I(f,\hat{n})R^{B*}_J(f,\hat{n}) + R^{L}_I(f,\hat{n})R^{L*}_J(f,\hat{n})\right]\,,
\end{aligned} \tag{8.81}$$

are the corresponding overlap functions for the tensor, vector, and scalar modes $\{T, V, S\}$. Note that $\Gamma^{(T)}_{IJ}(f)$ is identical to the ordinary overlap function $\Gamma_{IJ}(f)$ for an isotropic background (5.39).

Figure 65 show plots of the tensor, vector, and scalar overlap functions for the three different LIGO-Virgo detector pairs. The overlap functions have been normalized so that they equal 1 for colocated and coaligned detectors. This requires multiplying $\Gamma_{IJ}(f)$ by a factor of 5 for the tensor and vector overlap functions (5.43), but by a factor of 10 for the scalar overlap function.

8.4.3 Component separation: ML estimates of $S_h^{(T)}$, $S_h^{(V)}$, and $S_h^{(S)}$

Proceeding along the same lines as in Section 8.2.3, we now describe a method for separating the tensor, vector, and scalar contributions to the total strain spectral density. As shown in [130], we will need to analyze data from at least three independent baselines (so at least three detectors). As before, we will adopt the notation $\alpha = 1, 2, \cdots, N_b$ to denote the individual baselines (detector pairs) and α_1, α_2 to denote the two detectors that constitute that baseline.

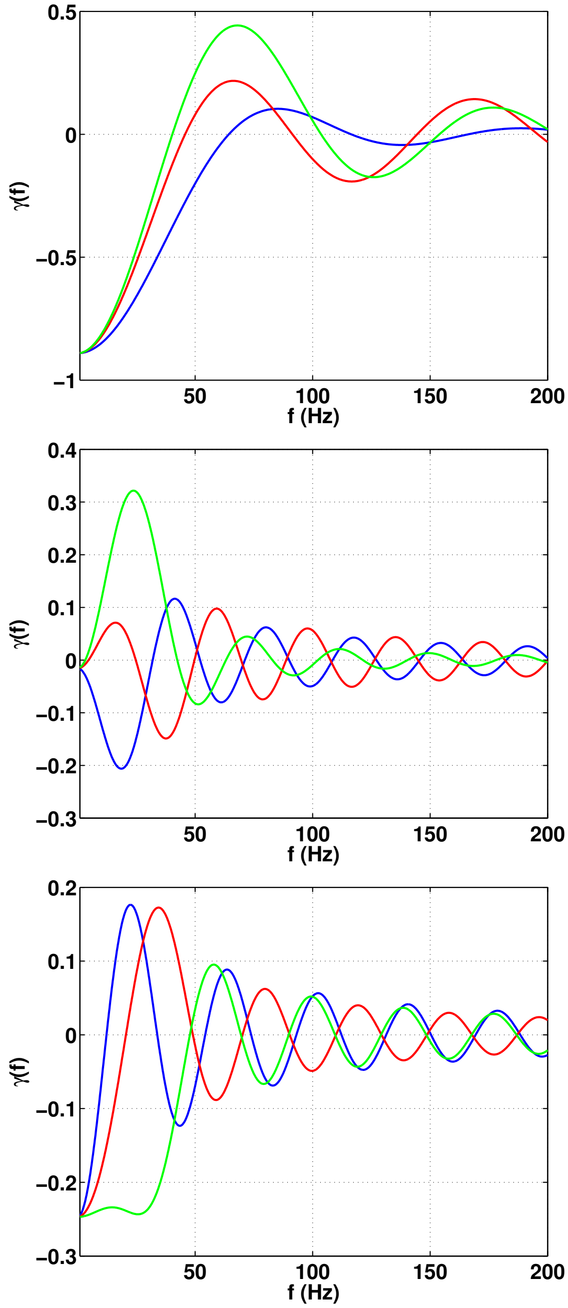


Figure 65: Normalized overlap functions for unpolarized tensor, vector, and scalar modes for the LIGO Hanford-LIGO Livingston detector pair (top panel); for the LIGO Hanford-Virgo detector pair (middle panel); and for the LIGO Livingston-Virgo detector pair (bottom panel). The tensor overlap functions are shown in blue; the vector overlap functions are shown in red; the scalar overlap functions are shown in green. These overlap functions were calculated in the small-antenna limit.

Our starting point is again the cross-correlated data from pairs of detectors in the network:

$$\hat{C}_\alpha(f) \equiv \frac{2}{T}\, \tilde{d}_{\alpha_1}(f)\tilde{d}^*_{\alpha_2}(f)\,, \tag{8.82}$$

where

$$\tilde{d}_{\alpha_I}(f) = \tilde{h}_{\alpha_I}(\tilde{f}) + \tilde{n}_{\alpha_I}(f)\,, \quad I = 1, 2\,. \tag{8.83}$$

Assuming that the noise in the individual detectors are uncorrelated with one another, it follows that

$$\langle \hat{C}_\alpha(f)\rangle = \Gamma^{(T)}_\alpha(f)S^{(T)}_h(f) + \Gamma^{(V)}_\alpha(f)S^{(V)}_h(f) + \Gamma^{(S)}_\alpha(f)S^{(S)}_h(f)\,. \tag{8.84}$$

In addition,

$$\begin{aligned}\mathcal{N}_{\alpha\alpha'}(f, f') &\equiv \langle \hat{C}_\alpha(f)\hat{C}^*_{\alpha'}(f')\rangle - \langle \hat{C}_\alpha(f)\rangle\langle \hat{C}^*_{\alpha'}(f')\rangle \\ &\approx \delta_{\alpha\alpha'}\delta_{ff'}P_{n_{\alpha_1}}(f)P_{n_{\alpha_2}}(f)\,,\end{aligned} \tag{8.85}$$

where $P_{n_{\alpha_I}}(f)$ are the one-sided power spectral densities of the noise in the detectors, and where we have assumed again that the gravitational-wave signal is weak compared to the detector noise. As we did in Section 8.2.3 we can write down a likelihood function for the cross-correlated data given the signal model (8.84):

$$p(\hat{C}|\mathcal{A}) \propto \exp\left[-\frac{1}{2}(\hat{C} - M\mathcal{A})^\dagger \mathcal{N}^{-1}(\hat{C} - M\mathcal{A})\right]. \tag{8.86}$$

Here we have adopted the matrix notation:

$$M \equiv \begin{bmatrix} \Gamma^{(T)}_1 & \Gamma^{(V)}_1 & \Gamma^{(S)}_1 \\ \Gamma^{(T)}_2 & \Gamma^{(V)}_2 & \Gamma^{(S)}_2 \\ \vdots & \vdots & \vdots \\ \Gamma^{(T)}_{N_b} & \Gamma^{(V)}_{N_b} & \Gamma^{(S)}_{N_b} \end{bmatrix}, \quad \mathcal{A} \equiv \begin{bmatrix} S^{(T)}_h \\ S^{(V)}_h \\ S^{(S)}_h \end{bmatrix}. \tag{8.87}$$

Since $\mathcal{A}$ enters quadratically in the exponential, we have the usual expression for the maximum-likehood estimators:

$$\hat{\mathcal{A}} = F^{-1}X\,, \tag{8.88}$$

where

$$F \equiv M^\dagger \mathcal{N}^{-1} M\,, \qquad X \equiv M^\dagger \mathcal{N}^{-1}\hat{C}\,, \tag{8.89}$$

with M and $\mathcal{N}$ given above, and with the standard proviso about possibly having to use singular-value decomposition to invert F. The uncertainty in the maximum-likelihood recoved values is given by the covariance matrix

$$\langle \hat{\mathcal{A}}\hat{\mathcal{A}}^\dagger\rangle - \langle \hat{\mathcal{A}}\rangle\langle \hat{\mathcal{A}}^\dagger\rangle \approx F^{-1}\,, \tag{8.90}$$

which we will use below to define *effective* overlap functions for the tensor, vector, and scalar modes for a multibaseline network of detectors.

8.4.4 Effective overlap functions for multiple baselines

For a multibaseline network of detectors, one has

$$F = \begin{bmatrix} \sum_\alpha N_\alpha^{-1}(\Gamma_\alpha^{(T)})^2 & \sum_\alpha N_\alpha^{-1}\Gamma_\alpha^{(T)}\Gamma_\alpha^{(V)} & \sum_\alpha N_\alpha^{-1}\Gamma_\alpha^{(T)}\Gamma_\alpha^{(S)} \\ \sum_\alpha N_\alpha^{-1}\Gamma_\alpha^{(V)}\Gamma_\alpha^{(T)} & \sum_\alpha N_\alpha^{-1}(\Gamma_\alpha^{(V)})^2 & \sum_\alpha N_\alpha^{-1}\Gamma_\alpha^{(V)}\Gamma_\alpha^{(S)} \\ \sum_\alpha N_\alpha^{-1}\Gamma_\alpha^{(S)}\Gamma_\alpha^{(T)} & \sum_\alpha N_\alpha^{-1}\Gamma_\alpha^{(S)}\Gamma_\alpha^{(V)} & \sum_\alpha N_\alpha^{-1}(\Gamma_\alpha^{(S)})^2 \end{bmatrix}, \tag{8.91}$$

where $N_\alpha(f) \equiv P_{n_{\alpha_1}}(f)P_{n_{\alpha_2}}(f)$. Let us assume that the determinant of the 3×3 matrices for each frequency (which we will denote by $\bar{F}$) are not equal to zero. Then the uncertainties in the estimators of $S_h^{(T)}$, $S_h^{(V)}$, and $S_h^{(S)}$ can be written as

$$\begin{aligned} \sigma_{\hat{T}}^2 &= (\bar{F}^{-1})_{TT} = \frac{1}{\det(\bar{F})}\left(\sum_\alpha N_\alpha^{-1}(\Gamma_\alpha^{(V)})^2 \sum_{\alpha'} N_{\alpha'}^{-1}(\Gamma_{\alpha'}^{(S)})^2 - \left(\sum_\alpha N_\alpha^{-1}\Gamma_\alpha^{(S)}\Gamma_\alpha^{(V)}\right)^2\right), \\ \sigma_{\hat{V}}^2 &= (\bar{F}^{-1})_{VV} = \frac{1}{\det(\bar{F})}\left(\sum_\alpha N_\alpha^{-1}(\Gamma_\alpha^{(T)})^2 \sum_{\alpha'} N_{\alpha'}^{-1}(\Gamma_{\alpha'}^{(S)})^2 - \left(\sum_\alpha N_\alpha^{-1}\Gamma_\alpha^{(S)}\Gamma_\alpha^{(T)}\right)^2\right), \\ \sigma_{\hat{S}}^2 &= (\bar{F}^{-1})_{SS} = \frac{1}{\det(\bar{F})}\left(\sum_\alpha N_\alpha^{-1}(\Gamma_\alpha^{(T)})^2 \sum_{\alpha'} N_{\alpha'}^{-1}(\Gamma_{\alpha'}^{(V)})^2 - \left(\sum_\alpha N_\alpha^{-1}\Gamma_\alpha^{(V)}\Gamma_\alpha^{(T)}\right)^2\right). \end{aligned} \tag{8.92}$$

Following [130], we can now define the *effective* overlap functions for the the tensor, vector, and scalar modes, associated with a multibaseline detector network. As we did in Section 8.2.5, we will assume for simplicity that the noise power spectra for the detectors are equal to one another so that $N_\alpha \equiv N$ can be factored out of the above expressions. We then define

$$\begin{aligned} \Gamma_{\rm eff}^{(T)}(f) &\equiv \sigma_{\hat{T}}^{-1}\sqrt{N} = \left(\frac{N^3\det(\bar{F})}{\sum_\alpha(\Gamma_\alpha^{(V)})^2\sum_{\alpha'}(\Gamma_{\alpha'}^{(S)})^2 - \left(\sum_\alpha \Gamma_\alpha^{(S)}\Gamma_\alpha^{(V)}\right)^2}\right)^{1/2}, \\ \Gamma_{\rm eff}^{(V)}(f) &\equiv \sigma_{\hat{V}}^{-1}\sqrt{N} = \left(\frac{N^3\det(\bar{F})}{\sum_\alpha(\Gamma_\alpha^{(T)})^2\sum_{\alpha'}(\Gamma_{\alpha'}^{(S)})^2 - \left(\sum_\alpha \Gamma_\alpha^{(S)}\Gamma_\alpha^{(T)}\right)^2}\right)^{1/2}, \\ \Gamma_{\rm eff}^{(S)}(f) &\equiv \sigma_{\hat{S}}^{-1}\sqrt{N} = \left(\frac{N^3\det(\bar{F})}{\sum_\alpha(\Gamma_\alpha^{(T)})^2\sum_{\alpha'}(\Gamma_{\alpha'}^{(V)})^2 - \left(\sum_\alpha \Gamma_\alpha^{(V)}\Gamma_\alpha^{(T)}\right)^2}\right)^{1/2}. \end{aligned} \tag{8.93}$$

Plots of $\Gamma_{\rm eff}^{(T)}(f)$, $\Gamma_{\rm eff}^{(V)}(f)$, and $\Gamma_{\rm eff}^{(S)}(f)$ are shown in Figure 66 for the multibaseline network formed from the LIGO Hanford, LIGO Livingston, and Virgo detectors. Dips in sensitivity correspond to frequencies where the determinant of $\bar{F}$ is zero (or close to zero).

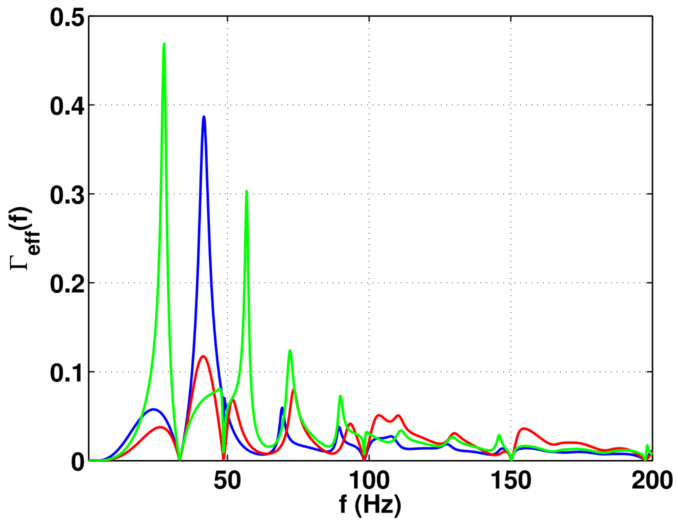


Figure 66: Effective overlap functions for $S_h^{(T)}$, $S_h^{(V)}$, $S_h^{(S)}$, for the multibaseline network formed from the LIGO Handford, LIGO Livingston, and Virgo detectors. $\Gamma_{\rm eff}^{(T)}(f)$ is shown in blue; $\Gamma_{\rm eff}^{(V)}(f)$ is shown in red; $\Gamma_{\rm eff}^{(S)}(f)$ is shown in green.

8.5 Searches for non-GR polarizations using pulsar timing arrays

As discussed in Section 8.3.4 it is also possible to search for non-GR polarizations using a pulsar timing array. Although the general concepts are the same as those for ground-based interferometers, there are some important differences, as the vector and scalar longitudinal polarization modes require keeping the pulsar term in the response functions to avoid possible singularities. We shall see below that the sensitivity to the vector and scalar longitudinal modes increases dramatically when cross-correlating data from pairs of pulsars with small angular separations. For additional details, see [109, 41, 71].

8.5.1 Polarization basis response functions

For pulsar timing, the response functions for Doppler frequency measurements for the different polarization modes $A = \{+, \times, X, Y, B, L\}$ are given by

$$R^A(f, \hat{n}) = \frac{1}{2}\frac{\hat{p}^a\hat{p}^b}{1-\hat{n}\cdot\hat{p}}e_{ab}^A(\hat{n})\left[1 - e^{-\frac{i2\pi fL}{c}(1-\hat{n}\cdot\hat{p})}\right], \tag{8.94}$$

where $\hat{p}$ points in the direction to the pulsar and L is its distance from Earth (see Section 5.2.1 with $\hat{p} = -\hat{u}$). Without loss of generality, we have assumed that the location of the measurement is at the origin of coordinates. Note that we have kept the *pulsar term* (the second term in the square brackets) since, as we shall see below, it is needed to get finite expressions for the response and overlap functions for the vector and scalar longitudinal modes.

Choosing our coordinate system so that $\hat{z}$ points along $\hat{p}$, we find:

$$
\begin{aligned}
R^{+}(f,\hat{n}) &= \frac{1}{2}(1+\cos\theta)\left[1-e^{-\frac{i2\pi fL}{c}(1-\cos\theta)}\right]\,,\\
R^{\times}(f,\hat{n}) &= 0\,,\\
R^{X}(f,\hat{n}) &= -\frac{\sin\theta\cos\theta}{1-\cos\theta}\left[1-e^{-\frac{i2\pi fL}{c}(1-\cos\theta)}\right]\,,\\
R^{Y}(f,\hat{n}) &= 0\,,\\
R^{B}(f,\hat{n}) &= \frac{1}{2}(1+\cos\theta)\left[1-e^{-\frac{i2\pi fL}{c}(1-\cos\theta)}\right]\,,\\
R^{L}(f,\hat{n}) &= \frac{1}{\sqrt{2}}\frac{\cos^2\theta}{1-\cos\theta}\left[1-e^{-\frac{i2\pi fL}{c}(1-\cos\theta)}\right]\,,
\end{aligned}
\tag{8.95}
$$

where we used (2.4) for our definitions of $\{\hat{n},\hat{l},\hat{m}\}$. Note that the response functions for the breathing mode B and the tensor $+$ mode have the same form for our particular choice of $\{\hat{l},\hat{m}\}$. This is not a problem, however, as we can still distinguish these modes due to their different behavior under rotations. The difference between the breathing and tensor modes becomes more apparent in terms of the spherical harmonic basis response functions $R^{B}_{(lm)}(f)$ and $R^{G}_{(lm)}(f)$, which are given in (8.98).

If we did not include the pulsar terms in the above expressions, then the response functions for both the vector and scalar longitudinal modes would become singular at $\theta=0$ (i.e., $\cos\theta=1$).[21] The factor of $\sin\theta$ in the numerator for $R^{X}(f,\hat{n})$ "softens" the $(1-\cos\theta)^{-1}$ singularity to $(1-\cos\theta)^{-1/2}$, so that it becomes integrable when calculating the vector longitudinal overlap functions [109, 41, 71]. (We will discuss this in more detail in Section 8.5.3.) By keeping the pulsar term we remove these singularities as can be seen by expanding the above expressions for $\theta\ll 1$:

$$
\begin{aligned}
R^{+}(f,\hat{n}) &\approx iy\theta^2/2\,,\\
R^{\times}(f,\hat{n}) &= 0\,,\\
R^{X}(f,\hat{n}) &\approx -iy\theta\,,\\
R^{Y}(f,\hat{n}) &= 0\,,\\
R^{B}(f,\hat{n}) &\approx iy\theta^2/2\,,\\
R^{L}(f,\hat{n}) &\approx iy/\sqrt{2}\,,
\end{aligned}
\tag{8.96}
$$

where $y\equiv 2\pi fL/c$, and we have assumed that $y\theta^2$ is also sufficiently small that we could Taylor expand the exponential. Since the typical distance to a pulsar is a few kiloparsecs and $f=3\times10^{-9}$ Hz for 10 yr of observation, we have $y\ \sim 10^4$, which means $\theta\lesssim 10^{-2}$ for the above expansions to be valid. Thus, for small angular separations between the direction to the pulsar and the direction to the gravitational wave, the reponse to the

[21] This corresponds to the direction to the pulsar and the direction to the source of the gravitational wave being the same. For this case, the radio pulse from the pulsar and the gravitational wave travel in phase with one another from the pulsar to Earth. It is as if the radio pulse "surfs" the gravitational wave [41].

scalar-longitudinal modes will be more than an order-of-magnitude larger than that for the vector modes, and several orders-of-magnitude larger than that for both the tensor and breathing modes. This increased sensitivity of the scalar longitudinal and vector longitudinal modes will also become apparent when we calculate the overlap functions for a pair of pulsars (see Section 8.5.3 and Figure 67).

8.5.2 Spherical harmonic basis response functions

It is also interesting to calculate the Doppler-frequency response functions for the tensor spherical harmonic components $P = \{G, C, V_G, V_C, B, L\}$. The general expression is given by:

$$R^P_{(lm)}(f) = \int d^2\Omega_{\hat{n}} \, \frac{1}{2} \frac{\hat{p}^a \hat{p}^b}{1 - \hat{n} \cdot \hat{p}} Y^P_{(lm)ab}(\hat{n}) \left[1 - e^{-\frac{i2\pi f L}{c}(1-\hat{n}\cdot\hat{p})}\right] . \tag{8.97}$$

As shown in [71], the above integral can be evaluated and then simplified by taking the limit $y \gg 1$, which as we mentioned above is valid for typical pulsars. The final results (taken from that paper) are:

$$\begin{aligned}
R^G_{(lm)}(f) &\approx 2\pi \, {}^{(2)}N_l Y_{lm}(\hat{p}) \, , && l = 2, 3, \cdots \, , \\
R^C_{(lm)}(f) &\approx 0 \, , && l = 2, 3, \cdots \, , \\
R^{V_G}_{(lm)}(f) &\approx 2\pi \, {}^{(1)}N_l \left[1 - \frac{2}{3}\delta_{l1}\right] Y_{lm}(\hat{p}) \, , && l = 1, 2, \cdots \, , \\
R^{V_C}_{(lm)}(f) &\approx 0 \, , && l = 1, 2, \cdots \, , \\
R^B_{(lm)}(f) &\approx 2\pi \, \frac{1}{\sqrt{2}} \left[\delta_{l0} + \frac{1}{3}\delta_{l1}\right] Y_{lm}(\hat{p}) \, , && l = 0, 1, \cdots \, , \\
R^L_{(lm)}(f) &\approx 2\pi \left[-\delta_{l0} - \frac{1}{3}\delta_{l1} + \frac{1}{2}\bar{H}_l(y)\right] Y_{lm}(\hat{p}) \, , && l = 0, 1, \cdots \, ,
\end{aligned} \tag{8.98}$$

where ${}^{(1)}N_l$ and ${}^{(2)}N_l$ are constants defined by (E.3) and (F.2), and

$$\bar{H}_l(y) \equiv \int_{-1}^{1} dx \, \frac{1}{(1-x)} P_l(x) \left(1 - e^{-iy(1-x)}\right) . \tag{8.99}$$

There are several important features to highlight about these expressions: (i) All of the response functions depend in the same way of the angular position of the pulsar, which is simply $Y_{lm}(\hat{p})$. (ii) Just as we saw earlier (5.24) that the response to the tensor curl mode is zero, so too is the response to the vector curl mode. *Thus, pulsar timing arrays are also insensitive to the curl component of the the vector-longitudinal modes.* (iii) In the limit $y \gg 1$, only the response to the scalar-longitudinal mode has frequency dependence (via y). (iv) The response to the breathing mode has non-zero contributions only from $l = 0$ and $l = 1$. In terms of power (which is effectively the square of the response), this means that pulsar timing observations will be insensitive to anisotropies in power in the breathing mode beyond quadrupole (i.e., $l = 2$).

8.5.3 Overlap functions

To calculate the overlap functions for non-GR polarization modes for pulsar timing arrays, we will proceed as we did in Section 8.4.2, assuming that the stochastic background is independently polarized, but is otherwise Gaussian-stationary and isotropic. (Extensions to *anisotropic* backgrounds will be briefly mentioned in Section 8.5.4. Details can be found in [71].) Making these assumptions, the quadratic expectation values of the Fourier coefficients $h_A(f,\hat{n})$ take the form

$$\langle h_A(f,\hat{n}) h^*_{A'}(f',\hat{n}')\rangle = \frac{1}{8\pi} S^A_h(f)\delta_{AA'}\delta(f-f')\delta^2(\hat{n},\hat{n}')\,, \tag{8.100}$$

where $S^A_h(f)$ are the one-sided strain spectral densities for the individual polarization modes. The overlap functions can then be calculated in the usual way, leading to

$$\langle \tilde{h}_I(f)\tilde{h}^*_J(f')\rangle = \frac{1}{2}\delta(f-f')\sum_A \Gamma^A_{IJ}(f) S^A_h(f)\,, \tag{8.101}$$

where

$$\Gamma^A_{IJ}(f) \equiv \frac{1}{4\pi}\int d^2\Omega_n R^A_I(f,\hat{n}) R^{A*}_J(f,\hat{n})\,. \tag{8.102}$$

Note the factor of $1/4\pi$ as compared to $1/8\pi$ in (8.81), and that there is no summation over A on the right-hand side of this expression.

For simplicity we will also assume as before that the tensor modes $\{+,\times\}$ and the vector-longitudinal modes $\{X,Y\}$ are unpolarized, so that

$$\begin{aligned} S^+_h(f) = S^\times_h(f) &\equiv \frac{1}{2} S^{(T)}(f)\,,\\ S^X_h(f) = S^Y_h(f) &\equiv \frac{1}{2} S^{(V)}(f)\,. \end{aligned} \tag{8.103}$$

Then we can define:

$$\begin{aligned} \Gamma^{(T)}_{IJ}(f) &\equiv \frac{1}{8\pi}\int d^2\Omega_{\hat{n}}\,\left[R^+_I(f,\hat{n})R^{+*}_J(f,\hat{n}) + R^\times_I(f,\hat{n})R^{\times *}_J(f,\hat{n})\right]\,,\\ \Gamma^{(V)}_{IJ}(f) &\equiv \frac{1}{8\pi}\int d^2\Omega_{\hat{n}}\,\left[R^X_I(f,\hat{n})R^{X*}_J(f,\hat{n}) + R^Y_I(f,\hat{n})R^{Y*}_J(f,\hat{n})\right]\,, \end{aligned} \tag{8.104}$$

for the unpolarized tensor and vector-longitudinal components. But we will keep the breathing and scalar-longitudinal overlap functions separate:

$$\begin{aligned} \Gamma^B_{IJ}(f) &\equiv \frac{1}{4\pi}\int d^2\Omega_{\hat{n}}\, R^B_I(f,\hat{n})R^{B*}_J(f,\hat{n})\,,\\ \Gamma^L_{IJ}(f) &\equiv \frac{1}{4\pi}\int d^2\Omega_{\hat{n}}\, R^L_I(f,\hat{n})R^{L*}_J(f,\hat{n})\,, \end{aligned} \tag{8.105}$$

given the complications that arise when trying to explicitly calculate $\Gamma^L_{IJ}(f)$.

As noted in Section 5.4.3, the overlap function for the tensor modes can be calculated analytically [82], without needing to include the pulsar term in the response functions:

$$\Gamma_{IJ}^{(T)} = \frac{1}{3}\left[\frac{3}{2}\left(\frac{1-\cos\zeta_{IJ}}{2}\right)\ln\left(\frac{1-\cos\zeta_{IJ}}{2}\right) - \frac{1}{4}\left(\frac{1-\cos\zeta_{IJ}}{2}\right) + \frac{1}{2}\right], \tag{8.106}$$

where ζ_{IJ} is the angle between two Earth-pulsar baselines, i.e., $\cos\zeta_{IJ} = \hat{p}_I \cdot \hat{p}_J$. The above expression differs from (5.55) by an overall normalization. The overlap functions for the breathing mode and for the vector longitudinal modes can be also be calculated analytically, again without needing to include the pulsar term in the reponse. For the breathing mode we have

$$\Gamma_{IJ}^{B} = \frac{2}{3}\left[\frac{3}{8} + \frac{1}{8}\cos\zeta_{IJ}\right]. \tag{8.107}$$

For the vector-longitudinal modes we have [109, 71]

$$\Gamma_{IJ}^{(V)} = \frac{1}{3}\left[\frac{3}{2}\ln\left(\frac{2}{1-\cos\zeta_{IJ}}\right) - 2\cos\zeta_{IJ} - \frac{3}{2}\right], \tag{8.108}$$

where we have assumed here that the angular separation ζ_{IJ} is not too small. In the limit $\zeta_{IJ} \to 0$, the above expression for $\Gamma_{IJ}^{(V)}$ diverges, which means that we need to include the pulsar terms in the response functions to handle that case. Doing so results in an expression that is finite, but depends on the frequency f via the distances to the pulsars, $2\pi f L_I/c$ and $2\pi f L_J/c$. (See Appendix J of [71] for an analytic expression for $\Gamma_{IJ}^{(V)}(f)$ in the limit $\zeta_{IJ} \to 0$.)

Finally, for the scalar longitudinal overlap function $\Gamma_{IJ}^{L}(f)$, there is no known analytic expression for the integral (8.105), except in the limit of codirectional ($\zeta_{IJ} = 0$) and anti-directional ($\zeta_{IJ} = \pi$) pulsars [109, 41, 71]. The pulsar terms need to be included in the scalar-longitudinal response functions for all cases to obtain a finite result, which again depends on the frequency f via the distances to the pulsars. A semi-analytic expression for $\Gamma_{IJ}^{L}(f)$ is derived in [71], which is valid in the $2\pi f L/c \gg 1$ limit. The semi-analytic expression effectively replaces the double integral over directions on the sky $\hat{n} = (\theta, \phi)$ with just a single numerical integration over θ. See [71] for additional details regarding that calculation.

Plots of the normalized overlap functions for the tensor, vector-longitudinal, breathing and scalar-longitudinal modes are shown in Figure 67, plotted as functions of the angular separation ζ between pairs of pulsars. The normalization is the same for each overlap function, chosen so that the tensor overlap function agrees with the normalized Hellings and Downs curve (5.55). The plots for the tensor, vector-longitudinal, and breathing modes are all real and do not depend on frequency; the plot for the scalar-longitudinal modes has both real and imaginary components (shown in red), and depends on frequency via the distances to the pulsars. For the scalar-longitudinal overlap function, we chose $y_1 = 1000$ and $y_2 = 2000$ for all pulsar pairs, where $y \equiv 2\pi f L/c$, and we did the integration numerically over both θ and ϕ. Note the different vertical scales for the vector-longitudinal and scalar-longitudinal overlap functions, compared to those for the tensor and breathing modes. For small angular separations, the sensitivity to vector-longitudinal modes is

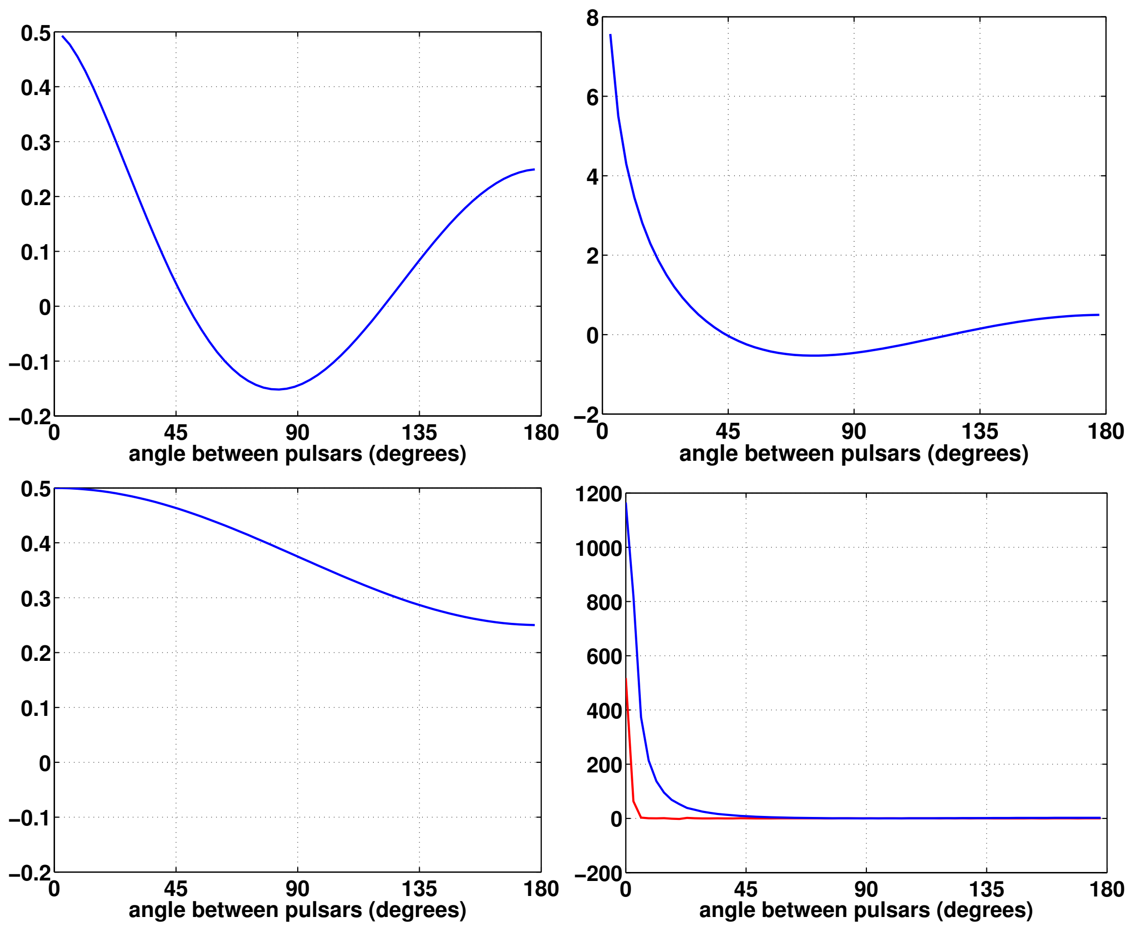


Figure 67: Normalized overlap functions for the tensor (upper left), vector-longitudinal (upper right), breathing (lower left), and scalar-longitudinal (lower right) polarization modes, plotted as functions of the angular separation between pairs of pulsars. The blue and red curves in the lower right-hand plot correspond to the real and imaginary parts of the scalar-longitudinal overlap function. Note the different vertical scales for the vector-longitudinal and scalar-longitudinal overlap functions, compared to those for the tensor and breathing modes.

roughly an order of magnitude larger than that for the tensor and breathing modes, while the sensitivity to the scalar-longitudinal mode is several orders-of-magnitude larger. This is consistent with what we found for the response functions, as discussed at the end of Section 8.5.1.

8.5.4 Component separation and anisotropic backgrounds

As shown in [71] the above calculations for non-GR polarization modes can be extended to *anisotropic* stochastic backgrounds. The spherical harmonic components of the overlap functions

$$\Gamma^A_{lm}(f) = \frac{1}{4\pi}\int d^2\Omega_{\hat{n}} Y_{lm}(\hat{n}) R^A_I(f,\hat{n}) R^{A*}_J(f,\hat{n}) \tag{8.109}$$

can be calculated *analytically* for the tensor, vector, and breathing polarization modes for all values of l and m, while the components of the scalar longitudinal overlap function admit only semi-analytic expressions. (This is similar to what we described in the previous section in the context of an isotropic background.) Plots of the first few spherical harmonics components, as a function of the angular separation ζ_{IJ} between a pair of pulsars, are given in Figures 1, 5, 2, and 3 of [71].

The ability to separate the contributions to the background from the different polarization modes depends crucially on the form of the spherical harmonic basis response functions $R^P_{(lm)}(f)$, where $P = \{G, C, V_G, V_C, B, L\}$. These were defined in (8.97) and have the $y \equiv 2\pi f L/c \gg 1$ limiting expressions given in (8.98). Recall that the (lm) indices here correspond to an expansion of the Fourier components of the metric perturbations in terms of tensor (spin 2), vector (spin 1), and scalar (spin 0) spherical harmonics:

$$h_{ab}(f,\hat{n}) = \sum_P \sum_{(lm)} a^P_{(lm)}(f) Y^P_{(lm)ab}(\hat{n})\,, \tag{8.110}$$

for which

$$\tilde{h}_I(f) = \sum_P \sum_{(lm)} R^P_{I(lm)}(f) a^P_{(lm)}(f) \tag{8.111}$$

is the response of pulsar I to the background. The expansion coefficients $a^P_{(lm)}(f)$ give the contributions of the different polarization modes to the background, and $R^P_{I(lm)}(f)$ are the response functions for those particular coefficients. For an angular resolution of order $180°/l_{\rm max}$, the total number of modes that are (in principle) accessible to a pulsar timing array with a sufficient number of pulsars is

$$N_m = 3(l_{\rm max}+1)^2 - 1\,. \tag{8.112}$$

This expression uses the result that the response to the curl modes for both the tensor and vector components are identically zero, as is the response to the breathing modes for $l \geq 2$. Since a pulsar timing array having N_p pulsars can measure at most $2N_p$ real components of the background (as discussed in Section 7.5.4), we see that at least $N_p = N_m$ pulsars are required to measure the N_m (complex) components.

But as noted in Section 8.5.2, all of the response functions $R^P_{(lm)}(f)$ depend on the direction $\hat{p}$ to the pulsar in exactly the same way, being proportional to $Y_{lm}(\hat{p})$. This degeneracy complicates the extraction of the different polarization modes. For the tensor and breathing modes, the degeneracy is broken since pulsar timing arrays typically operate in a regime where $y \gg 1$, for which the pulsar term can be ignored in the response functions for these modes. In that limit, a pulsar timing array is only sensitive to breathing modes with $l = 0, 1$, while the tensor modes are non-zero only for $l \geq 2$. On the other hand, the scalar-longitudinal and vector-longitudinal modes can only be distinguished from the tensor and breathing modes if there are multiple pulsars along the same line of sight, or if there is a known correlation between the expansion coefficients $a^P_{(lm)}(f)$ at different frequencies, e.g., a power-law spectrum. For either of these two cases, we can exploit the frequency dependence of the pulsar term, which is more significant for the longitudinal modes of the background. Keeping all of the frequency-dependent terms [71]:

$$\begin{aligned} R^L_{lm}(f) = 2\pi(-1)^l \bigg\{ & -\delta_{l0} + \frac{1}{3}\delta_{l1} \\ & + (-i)^l \mathrm{e}^{-iy} \left[\left(1 - i\frac{l}{y} \right) j_l(y) + i j_{l+1}(y) \right] + \frac{1}{2} H_l(y) \bigg\} Y_{lm}(\hat{p}) \,, \end{aligned} \tag{8.113}$$

and

$$\begin{aligned} R^{V_G}_{lm}(f) = \pi(-1)^l \, {}^{(1)}N_l \bigg\{ & \frac{4}{3}\delta_{l1} + 2(-i)^l \mathrm{e}^{-iy} \bigg[\left(1 - \frac{il}{y} \right) (l+1) j_l(y) \\ & - (y - i(2l+3)) j_{l+1}(y) - iy j_{l+2}(y) \bigg] \bigg\} Y_{lm}(\hat{p}) \,, \end{aligned} \tag{8.114}$$

for the scalar-longitudinal and vector-longitudinal response functions, where $j_l(y)$ denotes a spherical Bessel functions of order l and $y \equiv 2\pi f L/c$. If we take the $y \gg 1$ limit of these equations, we recover the approximate expressions given in (8.98). But to separate the various components of the background, we need to use these more complicated expressions to break the angular-direction degeneracy.

A quantitative analysis of the sensitivity of a phase-coherent mapping search (Section 7.5) to the different components $a^P_{(lm)}(f)$ of a stochastic background is given in [71]. The results of that analysis are summarized in Table 8.5.4, which is taken from that paper. The entries in the table show how the uncertainties in our measurements change as we search for: (i) only the tensor modes, (ii) both tensor and breathing modes, (iii) tensor, breathing, and scalar-longitudinal modes, and (iv) all possible modes. The uncertainties were obtained by taking the square root of the diagonal elements of the inverse of the Fisher matrix, following the general prescription described in Section 7.5.1. For this calculation, 30 pulsars were distributed randomly on the sky, with distances chosen at random, uniformly between 1 and 10 kpc. There was only a single frequency component, $f_0 = 3 \times 10^{-9}$ Hz, and the timing noise was assumed to be the same for all the pulsars in the array. The background was also assumed to contain modes with equal intrinsic amplitudes up to $l_{\max} = 2$, so that the total number of modes $N_m = 26$ was less than the number of pulsars in the array. This gave a fully-determined system of equations that needed to be solved.

	(l,m) mode								
	$(0,0)$	$(1,-1)$	$(1,0)$	$(1,1)$	$(2,-2)$	$(2,-1)$	$(2,0)$	$(2,1)$	$(2,2)$
tensor	–	–	–	–	0.44	0.38	0.32	0.38	0.44
tensor	–	–	–	–	0.49	0.39	0.37	0.39	0.49
breathing	0.16	0.53	0.46	0.53	–	–	–	–	–
tensor	–	–	–	–	16.2	10.5	11.4	10.5	16.2
breathing	4.36	16.1	14.1	16.1	–	–	–	–	–
longitudinal	0.71	0.96	0.84	0.96	1.21	0.78	0.86	0.78	1.21
tensor	–	–	–	–	1.4e5	5.4e4	8.0e4	5.4e4	1.4e5
breathing	18.4	9.4e4	6.2e4	9.4e4	–	–	–	–	–
longitudinal	3.08	11.5	8.68	11.5	20.9	7.51	11.9	7.52	20.9
vector	–	6.6e4	4.4e4	6.6e4	7.0e4	2.7e4	4.0e4	2.7e4	7.0e4

Table 7: Relative uncertainties for the tensor, breathing, scalar-longitudinal, and vector-longitudinal polarization modes searched for separately or in various combinations for $l_{\max} = 2$ and $N_p = 30$ pulsars. This table is adapted from Table II in [71].

The entries in the table reflect our expectations for recovering the different modes of the background. Namely, there is little change in our ability to recover the tensor modes when the breathing modes are also included in the analysis. This is because the tensor modes are non-zero only for $l \geq 2$, while the response to the breathing modes is non-zero only for $l = 0, 1$. Adding the scalar-longitudinal modes to the analysis worsens the recovery of the tensor and breathing modes by about an order of magnitude, as the scalar-longitudinal modes can also have non-zero values for all values of l. (There are simply more parameters to recover.) But one is still able to break the degeneracy as the response to the scalar-longitudinal modes depends *strongly* on the distances to the pulsars. The uncertainity in the recovery of the scalar-longitudinal modes is about an order of magnitude less than that for the tensor and breathing modes, since the analysis assumes equal intrinsic amplitudes for all the modes, while the correlated response to the scalar-longitudinal modes is much larger for small angular separations between the pulsars (Section 8.5.3 and Figure 67). Finally, adding the vector-longitudinal modes to the analysis weakens the recovery of the scalar-longitudinal modes by about an order of magnitude, again because more parameters need to be recovered. However, it *severely worsens* the recovery of all the other modes, because of the degeneracy in the response on the angular direction to the pulsars. There is some dependence on frequency for the vector-longitudinal response, as indicated in (8.114), but it is much weaker than the frequency dependence of the scalar-longitudinal modes. So the degeneracy is not broken nearly as strongly for these modes. See [71] for more details.

8.6 Other searches

It is also possible to use the general cross-correlation techniques described in Section 4 to search for signals that don't really constitute a stochastic gravitational-wave background. Using a stochastic-based cross-correlation method to search for such signals is not optimal, but it still gives valid results for detection statistics or estimators of signal parameters, with error bars that properly reflect the uncertainty in these quantities. It is just that these error bars are *larger* than those for an optimal (minimum variance) search, which is better "tuned" for the signal. Below we briefly describe how the general cross-correlation method can be used to search for (i) long-duration unmodelled transients and (ii) persistent (or continuous) gravitational waves from targeted sources.

8.6.1 Searches for long-duration unmodelled transients

The Stochastic Transient Analysis Multi-detector Pipeline [179] (STAMP for short) is a cross-correlation search for unmodelled long-duration transient signals ("bursts") that last on order a few seconds to several hours or longer. The duration of these transients are long compared to the typical merger signal from inspiralling binaries (tens of milliseconds to a few seconds), but short compared to the persistent quasi-monochromatic signals that one expects from e.g., rotating (non-axisymmetric) neutron stars. STAMP was developed in the context of ground-based interferometers, but the general method, which we briefly describe below, is also valid for other types of gravitational-wave detectors.

STAMP is effectively an adapted gravitational-wave radiometer search (Section 7.3.6), which cross-correlates data from pairs of detectors (7.6), weighted by the *inverse* of the integrand of the overlap function $\gamma_{IJ}(t; f, \hat{n})$ for a particular direction $\hat{n}$ on the sky:

$$\tilde{s}_{IJ}(t, f; \hat{n}) \equiv \frac{2}{\tau} \frac{\tilde{d}_I(t; f)\tilde{d}_J^*(t; f)}{\gamma_{IJ}(t; f, \hat{n})} \,, \tag{8.115}$$

where τ is the duration of the segments defining the short-term Fourier transforms. The weighting by the inverse of $\gamma_{IJ}(t; f, \hat{n})$ is used so that the expected value of $\tilde{s}_{IJ}(t; f, \hat{n})$ is just the gravitational-wave power in pixel $(t; f)$ for a point source in direction $\hat{n}$, which follows from (7.7). The data $\tilde{s}_{IJ}(t, f; \hat{n})$ for a single direction $\hat{n}$ define a *frequency-time map*. For a typical analysis using the LIGO Hanford and LIGO Livingston interferometer, a single map has a frequency range from about 50 to $\sim$1000 Hz, and a time duration of a couple hundred seconds (or whatever the expected duration of the transient might be). A strong burst signal shows up as *cluster* or *track* of bright pixels in the frequency-time map, which stands out above the noise. The data analysis problem thus becomes a *pattern recognition* problem.

The procedure for deciding whether or not a signal is present in the data can be broken down into three steps: (i) determine if a statistically significant clump or track of bright pixels is present in a frequency-time map, which requires using some form of pattern-recognition or clustering algorithm (see [179] and relevant references cited therein); (ii) calculate the value of the detection statistic Λ, obtained from a weighted sum of the power in the pixels for each cluster determined by the previous step; (iii) compare the observed value of the detection statistic to a threshold value Λ_*, which depends on the desired false alarm rate. This threshold is typically calculated by time-shifting the data to empirically determine the sampling distribution of Λ in the absence of a signal. If $\Lambda_{\text{obs}} > \Lambda_*$, then reject the null hypothesis and claim detection as discussed in 3.2.1. (Actually, in practice, this last step is a bit more complicated, as one typically does follow-up investigations using auxiliary instrumental and environmental channels, and data quality indicators. This provides additional confidence that the gravitational-wave candidate is not some spurious instrumental or environmental artefact.)

Figure 68 is an example of a frequency-time map with a simulated long-duration gravitational-wave signal injected into simulated initial LIGO detector noise. This particular signal is an *accretion disk instability waveform*, based on a model by van Putten [192, 194, 193]. The signal is a (inverse) "chirp" in gravitational radiation having an exponentially decaying frequency. (The magnitude of the signal increases with time as the frequency *decreases*.) The injected signal is strong enough to be seen by eye in the raw frequency-time map (left panel). After applying a clustering algorithm, the fluctuations in the detector noise have been noticeably reduced (right panel).

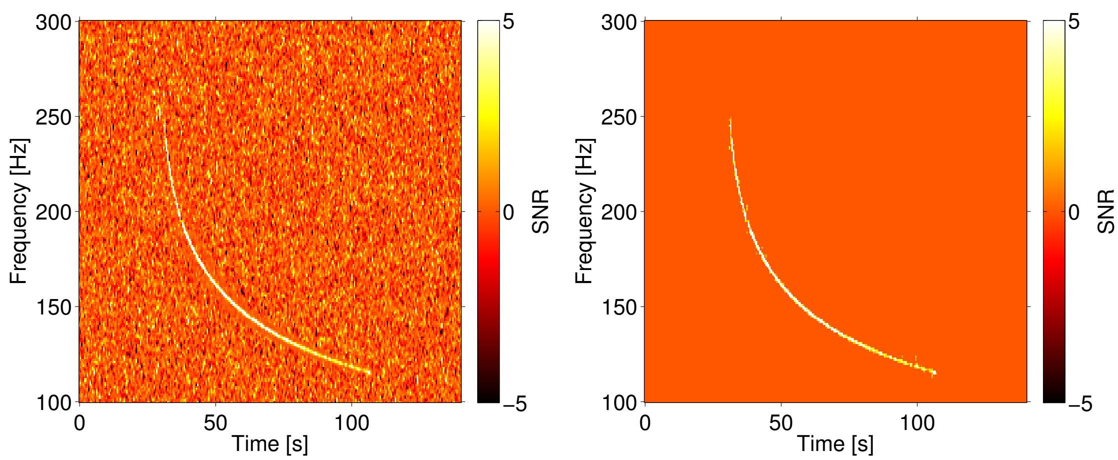


Figure 68: Frequency-time maps for an injected long-duration transient gravitational-wave signal in noise. Left panel: signal-to-noise ratio map before processing. Right panel: signal-to-noise ratio map after applying a clustering algorithm. Note that the noise fluctuations have effectively been eliminated in the second plot. (Figures provided by Tanner Prestegard.)

Readers should see [179] for many more details regarding STAMP, and [9, 1] for results from analyses of LIGO data taken during their 5th and 6th science runs—the first paper describes an all-sky search for long-duration gravitational-wave transients; the second, a triggered-search for long-duration gravitational-transients coincident with long duration gamma-ray bursts.

8.6.2 Searches for targeted-sources of continuous gravitational waves

The gravitational-wave radiometer method (Section 7.3.6) can also be used to look for gravitational waves from persistent (continuous) sources at known locations on the sky, e.g., the galactic center, the location of SN 1987A, or from low-mass X-ray binaries like Sco X-1 [5, 117]. For example, Sco X-1 is expected to emit gravitational waves from the (suspected) rotating neutron star at its core, having non-axisymmetric distortions produced by the accretion of matter from the low-mass companion. The parameters of this system that determine the phase evolution of the gravitational radiation are not well-constrained: (i) Since the neutron star at the core has not been observed to emit pulsations in the radio or any electromagnetic band, the orbital parameters of the binary are estimated instead from optical observations of the low-mass companion [163, 73]. These observations do not constrain the orbital parameters as tightly as being able to directly monitor the spin frequency of the neutron star. (ii) The intrinsic spin evolution of the neutron star also has large uncertainties due to the high rate of accretion of from the low-mass companion star. Both of these features translate into a *large* parameter space volume over which to search, making fully-coherent matched-filter searches for the gravitational-wave signal computationally challenging [117].

Nonetheless, for such sources, one can perform a *narrow-band, targeted* radiometer search, cross-correlating data from a pair of detectors with a filter function proportional to the integrand $\gamma_{IJ}(t; f, \hat{n}_0)$ of the overlap function evaluated at the direction $\hat{n}_0$ to the source on the sky:

$$\hat{C}_{IJ}(f) = \frac{2}{\tau} \sum_t \tilde{d}_I(t; f) \tilde{d}_J^*(t; f) \frac{\mathcal{N}_{IJ}(f) \gamma_{IJ}(t; f, \hat{n}_0)}{P_{n_I}(t; f) P_{n_J}(t; f)}, \tag{8.116}$$

where

$$\mathcal{N}_{IJ}(f) \equiv \left[\sum_t \frac{\gamma_{IJ}^2(t; f, \hat{n}_0)}{P_{n_I}(t; f) P_{n_J}(t; f)} \right]^{-1}. \tag{8.117}$$

The search is narrow-band in the sense that one doesn't integrate over the whole frequency band of the detectors, but looks instead for evidence of a gravitational wave in narrow frequency bins that span the sensitive band of the detector. The weighted cross-correlations are summed over time, to build up signal-to-noise ratio, since the source is assumed to be persistent. The frequentist detection statistic is the (squared) signal-to-noise ratio of the cross-correlated power contained in each narrow frequency band:

$$\Lambda(d) = \frac{|\hat{C}_{IJ}(f)|^2}{\mathrm{Var}[\hat{C}_{IJ}(f)]} \approx \frac{|\hat{C}_{IJ}(f)|^2}{\mathcal{N}_{IJ}(f)}, \tag{8.118}$$

where we used the result that the variance of the cross-correlation estimator $\hat{C}_{IJ}(f)$ equals the normalization factor $\mathcal{N}_{IJ}(f)$ in the weak-signal limit. This modified radiometer search is *robust* in the sense that it makes minimal assumptions about the source. The detection efficiency of the search could be improved if one had additional information about the signal (e.g., if one knew that the radiation was circularly polarized), which could then be included in the stochastic signal model.

9 Real-world complications

> Experience with real-world data, however, soon convinces one that both stationarity and Gaussianity are fairy tales invented for the amusement of undergraduates. *D.J. Thompson* [173]

The analyses described in the previous sections assumed that the instrument noise is stationary, Gaussian distributed, and uncorrelated between detectors. The analyses also implicitly assumed that the data were regularly sampled and devoid of gaps, facilitating an easy transition between the recorded time series and the frequency domain where many of the analyses are performed. In practice, all of these assumptions are violated to varying degrees, and the analyses of real data require additional care. Analyses that assume stationary, Gaussian noise can produced biased results when applied to more complicated real-world data sets.

Analysis of data from the first and second generation ground-based interferometers have shown that the data are neither perfectly stationary nor Gaussian. The non-stationarity can be broadly categorized as having two components: slow, adiabatic drifts in the noise spectrum with time; and short-duration noise transients, referred to as *glitches*, that have compact support in time-frequency. These glitches are also the dominant cause of non-Gaussianity in the noise distributions, giving rise to long tails that extend past a core distribution that is well described as Gaussian. The data are evenly sampled by design, though there are often large gaps between data segments due to loss of lock, scheduled maintenance, etc.

Pulsar timing data are, in many ways, far more challenging to analyze. The lack of dedicated telescope facilities, and the practical constraints associated with making the observations, result in data that are irregularly sampled. Moreover, the very long observation timelines (years to decades) and the mixture of facilities yield data sets that have been collected using a variety of receivers, data recorders, and pulse folding schemes. The heterogeneity of the observations causes the data to be non-stationary. An additional complication is that a complicated deterministic timing model that predicts the time of arrival of each pulse has to be subtracted from the data to produce the timing residuals used in the gravitational-wave analyses. The timing model includes a pulsar spindown model and a detailed pulse propagation model that accounts for the relative motion of the Earth and pulsar. Many of the pulsars are in binary systems, so the timing model has to include relativistic orbital motion, and propagation effects such as the Shapiro time delay.

For future space detectors we can only guess at the nature of the noise. Results from the LISA Pathfinder mission provide some insight [29], but only for a subset of the detector components, and for somewhat different flight hardware. The data will be regularly sampled, but data gaps are expected due to re-pointing of the communication antennae and orbit adjustments. Possible sources of non-stationarity include variations in the solar wind, thermal variations, and tidal perturbations from the Earth and other solar system bodies. The plans for the first space interferometers envision a single array of 3 spacecraft with 6 laser links. From these links three noise-orthogonal signal channels can be synthesized, but these combinations are also signal orthogonal, and so cross-correlation cannot be used to detect a signal.

9.1 Non-stationary noise

Data from existing gravitational-wave detectors, including bars, interferometers, and pulsar timing, exhibit various degrees of non-stationarity. Here we give examples relevant to ground based interferometers, but the situation is similar for the other detection techniques.

Non-stationary behavior can manifest itself in many forms, and there are no doubt many factors that contribute to the non-stationarity seen in interferometer data. Nonetheless, a simple two-part model does a good job of capturing the bulk of the non-stationary features. The two-part model consists of a slowly varying noise spectral density $S_n(t; f)$, and localized noise transients or "glitches". The slow drift in the spectrum can be modeled as a locally stationary noise process [55], which has the nice feature that for small enough time segments, the data in each segment can be treated as stationary. The glitch contribution to the non-stationarity poses more of a challenge, as the non-stationarity persists even for short data segments.

9.1.1 Local stationarity

To illustrate the two-component description of non-stationary data, we begin with a toy model of a locally stationary red noise process. Later we will add a model for the impulsive, glitch component (Section 9.1.2). Consider an auto-regressive AR(1) process of the form:

$$x_t = q(t)x_{t-1} + \epsilon(t)\delta\,, \tag{9.1}$$

where $\delta \sim N(0,1)$ is a unit Gaussian deviate and $q(t)$ and $\epsilon(t)$ are slowly-varying functions of time. The local power spectrum $S(t; f)$ for this process has the form

$$S(t;f) = \frac{\epsilon^2(t)}{1 + q^2(t) - 2q(t)\cos(\pi f/f_N)}\,, \tag{9.2}$$

where f_N is the Nyquist frequency. For a data segment of duration T with N samples, $f_N = N/(2T)$. Figure 69 shows the average and local spectra for $T = 1024$ seconds of data sampled at 1024 Hz with

$$q(t) = q_0 \frac{(1 + \alpha\cos(2\pi t/T))}{(1+\alpha)}\,, \qquad \epsilon(t) = \epsilon_0(1 + t/T)\,, \tag{9.3}$$

and $q_0 = 0.95$, $\alpha = 0.4$, and $\epsilon_0 = 1$. The local spectra are computed using 32-second segments of data that are smoothed and compared to the predicted spectra (9.2). The smoothed average spectrum is computed using the full data set and compared to the theoretical average spectrum

$$S(f) = \frac{1}{T}\int_0^T S(t; f)\, dt\,. \tag{9.4}$$

The high degree of non-stationarity is clearly apparent from the several orders of magnitude variation in the spectra across different segments of data. In LIGO stochastic background analyses a "delta sigma" cut is used to discard segments of data that exhibit significant

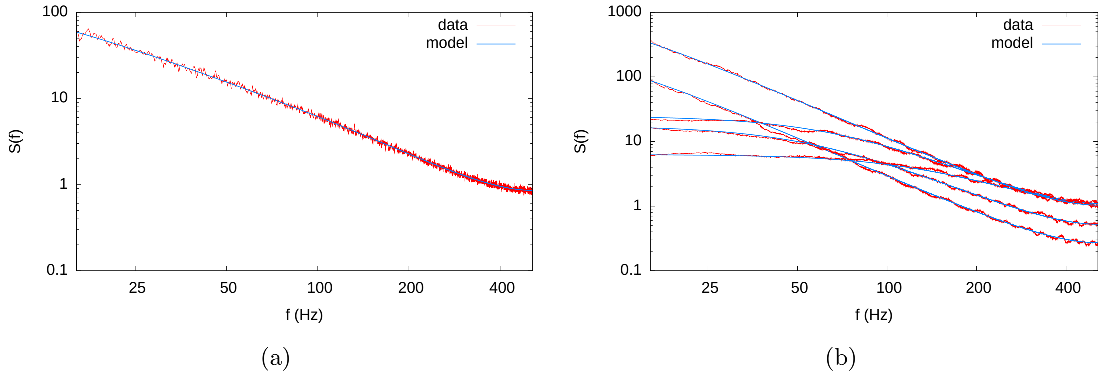


Figure 69: Spectra for the locally stationary AR(1) model. Panel (a) shows the smoothed spectrum computed using the full data set compared to the time average of the theoretical spectrum. Panel (b) shows smoothed spectra from the $1^{\rm st}$, $8^{\rm th}$, $16^{\rm th}$, $24^{\rm th}$ and $32^{\rm nd}$ time segments compared to the theoretical $S(t;f)$ computed at the central time for each segment.

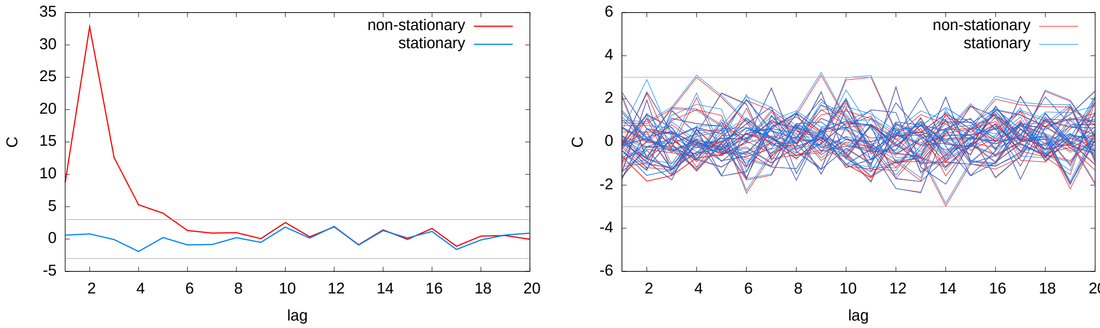
non-stationarity. The square-root of the variance (6.32) of the cross-correlation statistic is compared between three consecutive short segments of data (typically each 60 seconds long), and if the levels differ by more than 20%–30% those segments are not used in the analysis [7, 8].

The degree of non-stationarity can be measured from the auto-correlation of the whitened Fourier coefficients $\bar{x}_f = \tilde{x}_f/\sqrt{\hat{S}(f)}$, where $\hat{S}(f)$ is estimated from the smoothed power spectra. The auto-correlation at lag k is defined by

$$c(k) \equiv \frac{1}{2N}\sum_{i=1}^{N}(\bar{x}_i\bar{x}^*_{i+k} + \bar{x}^*_i\bar{x}_{i+k})\,. \tag{9.5}$$

For stationary, Gaussian noise in the large-N limit, $c(k)$ for $k > 0$ is Gaussian distributed with zero mean and variance $\sigma^2 = 1/N$ [59]. It is convenient to use the scaled auto-covariance $C(k) \equiv \sqrt{N}c(k)$, which has unit variance for stationary, Gaussian noise. Figure 70 compares $C(k)$ computed for the locally stationary $AR(1)$ model shown in Figure 69, and a stationary $AR(1)$ model with $q(t) = q_0$ and $\epsilon(t) = \epsilon_0$. The locally stationary model shows clear departures from stationarity when the auto-correlation is computed using the full data set (as evidenced by the large autocorrelations for small lags), while the data in each of the 32 sub-segments shows no signs of non-stationarity.

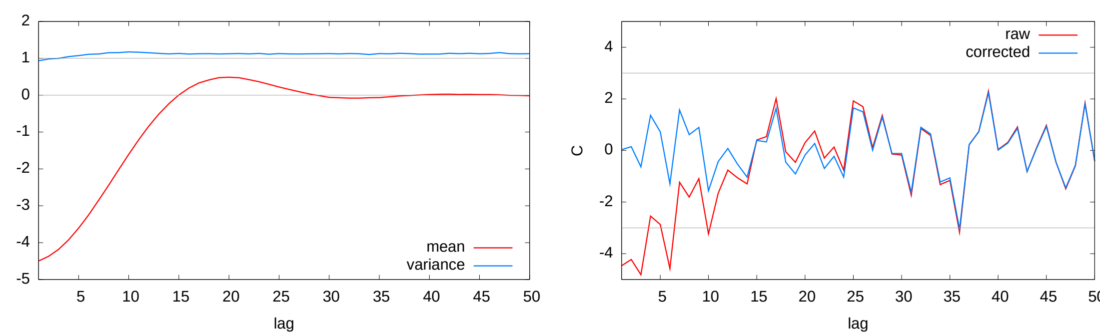
One note of caution in using the Fourier autocorrelation $C(k)$ as an indicator of non-stationarity is that any window that is applied to the time-domain data to lessen spectral leakage in the Fourier transform necessarily makes the data non-stationary. Choosing a window function that is unity across most of the samples, such as a Tukey window or a Plank taper window, lessens the taper-induced non-stationarity, but does not eliminate the effect. The solution is to apply a correction to the autocorrelation that accounts for



(a) (b)

Figure 70: Autocorrelation of the whitened Fourier coefficients as a function of lag for stationary and locally-stationary AR(1) models. Panel (a) is a comparison for the full data sets, and panel (b) is for each of the 32 sub-segments. The locally-stationary data show clear departures from stationarity in the full data set, but are consistent with stationarity in the shorter sub-segments of data.

the window. Figure 71 shows the impact that a Tukey window has on the mean and variance of the Fourier autocorrelation $C(k)$. In this simulation $N = 32768$ samples were used with a Tukey window that is constant across the central 90% of the samples. By subtracting the mean and scaling by the square-root of the variance caused by the Tukey window, the non-stationarity caused by the filter can be corrected for.



(a) (b)

Figure 71: Panel (a) shows the mean and variance of the autocorrelation for stationary, Gaussian noise caused by a Tukey window. Panel (b) shows the raw and corrected autocorrelation for a stationary, Gaussian noise process.

9.1.2 Glitches

To model the second form of non-stationarity caused by short-duration noise transients, we add Gaussian-enveloped noise bursts to stationary $AR(1)$ data. The bursts are simulated by generating white noise in the time domain, that is then multiplied by a Gaussian window centered at time t_0 with width σ_t. The data is then Fourier transformed, and the Fourier coefficients are multiplied by a Gaussian window centered at f_0 with width σ_f. In the simulation, the central times were drawn from a Poisson process with a rate of 0.5 Hz, and the central frequencies were drawn from a uniform distribution $U[0, f_N]$. The duration and bandwidth were also drawn from uniform distributions: $\sigma_t \sim U[0.01\,\mathrm{s}, 0.05\,\mathrm{s}]$, $\sigma_f \sim U[2\,\mathrm{Hz}, 50\,\mathrm{Hz}]$. The signal-to-noise ratio of the bursts was drawn from the distribution

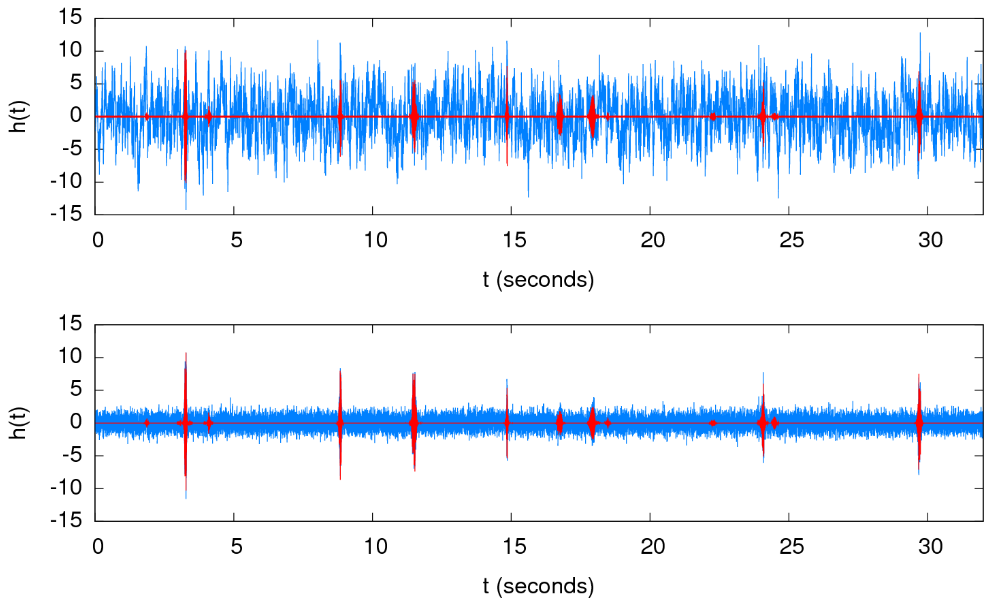
$$p(\mathrm{SNR}) = \frac{\mathrm{SNR}}{2\,\mathrm{SNR}_*^2 \left(1 + \frac{\mathrm{SNR}}{2\,\mathrm{SNR}_*}\right)^3} \,. \tag{9.6}$$

This form for the SNR distribution is used by the BayesWave algorithm [48] as a prior on the amplitude of glitches. The truncated power-law form for $p(\mathrm{SNR})$ is motivated by the distribution of glitches seen in real data. Figure 72 shows a 32-second segment of simulated data, and the dramatic effect that the glitches have on the autocorrelation of the Fourier transform. Unlike the locally stationary noise process, which only introduced correlations for small lags, the glitches produce a much larger deviation from stationarity that extends to large lags.

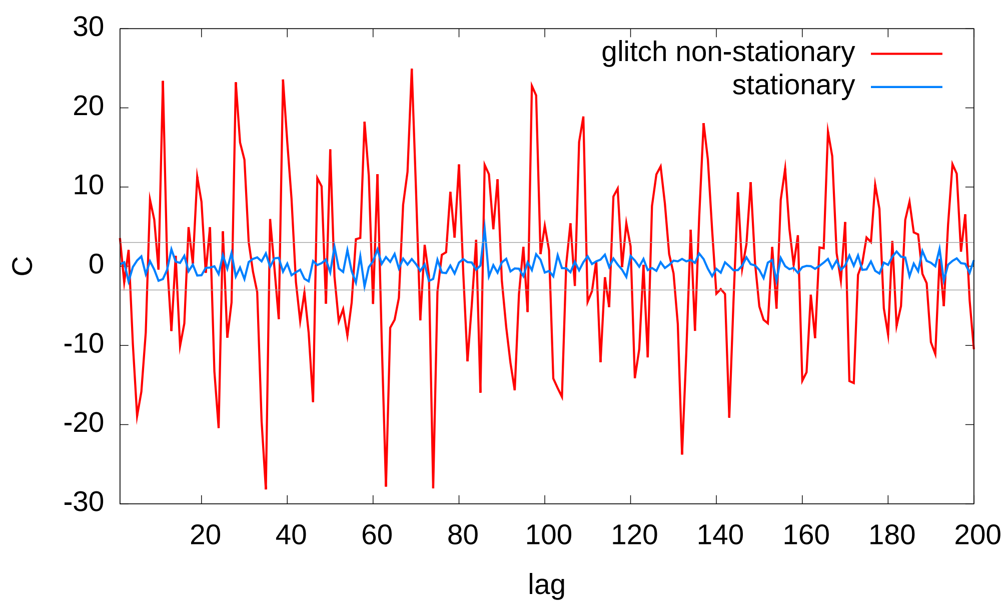
9.2 Non-Gaussian noise

Gaussian noise processes are ubiquitous in nature, and provide a remarkably good model for the data seen in gravitational-wave detectors. Properly whitened gravitational-wave data typically have a Gaussian core that accounts for the bulk of the samples, along with a small number of outliers in the tails of the distribution. Even these small departures can severely impact analyses that assume perfectly Gaussian distributions.

Gauss developed the *least-squares* (maximum-likelihood) data analysis technique in an effort to determine the orbit of the newly discovered dwarf planet Ceres. Gauss showed that if measurement errors are: (i) more likely small than large, (ii) symmetric, and (iii) have zero mean, then they follow a normal distribution (first described by de Moivre in 1733). Gauss' proof relied on the law of large numbers: he assumed that under repeated measurements the most-likely value of a quantity is given by the mean of the measured values. The assumptions used in Gauss' derivation were placed on a firmer footing by Laplace, who derived the central limit theorem, which states that the arithmetic mean of a sufficiently large number of independent random deviates will be approximately normally distributed, regardless of the underlying distributions the deviates are drawn from, so long as the distributions have finite first and second moments. The central limit theorem is often invoked to explain the ubiquity of Gaussian measurement errors. While the classic central limit theorem applies to noise contributions that are fundamentally stochastic (such as those with a quantum origin), a variant of the central limit theorem also applies to the sum of a large number of *deterministic* effects, so long as the deterministic processes obey certain conditions [91].



(a)



(b)

Figure 72: Panel (a) shows simulated stationary $AR(1)$ data with non-stationary noise transients, or glitches, highlighted in red. The upper panel is the raw data, while the lower panel has been whitened by the estimated amplitude spectral density. Panel (b) shows the autocorrelation for the stationary $AR(1)$ data without glitches in blue, and with glitches in red.

Since gravitational-wave data typically have highly colored spectra, one cannot simply compare the distribution of samples in time or frequency to a Gaussian distribution. The data first have to be whitened. This can be done by dividing the Fourier coefficients by an estimate of the power spectra and inverse Fourier transforming the result to arrive at a whitened time series. Figure 73 shows histograms of the whitened Fourier-domain and time-domain samples for the simulated data shown in Figure 72. By eye, the

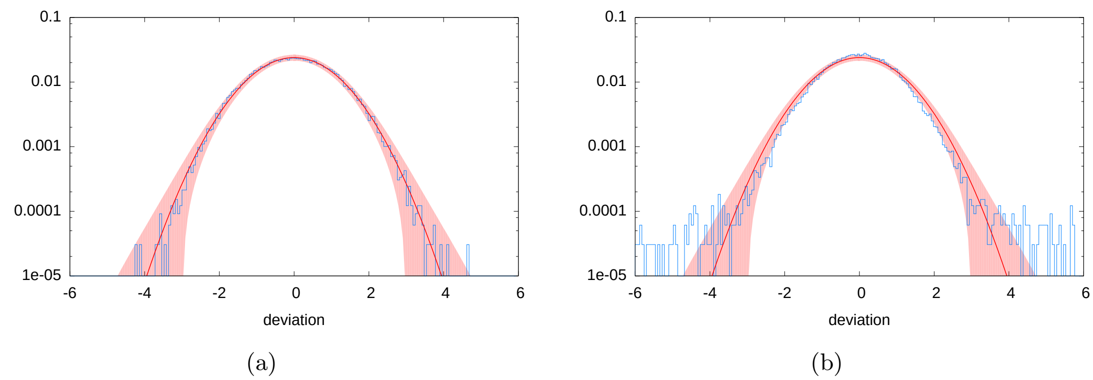


Figure 73: Histograms of the whitened data samples for the simulated data shown in Figure 72. A reference $N(0,1)$ Gaussian distribution is shown as a red line. The light red band denotes the 3-sigma confidence interval for the finite number of samples used to produce the histograms. Panel (a) uses the whitened Fourier coefficients, while panel (b) uses the whitened time-domain samples. While the non-Gaussianity is most apparent in the time domain, both distributions fail the Anderson-Darling test for Gaussianity.

frequency-domain samples appear fairly Gaussian, while the time-domain samples show clear departures from Gaussianity. Applying the Anderson-Darling test [27] to both sets of samples indicates that the Gaussian hypothesis is rejected in both cases, with a p-value of $p = 2.6 \times 10^{-5}$ for the Fourier-domain samples and $p < 10^{-20}$ for the time-domain samples. Applying the same analysis to the locally stationary $AR(1)$ model generated using 32 seconds of data (i.e., setting $T = 32$ s in the model for $q(t)$ and $\epsilon(t)$), we find that the whitened Fourier coefficients generally pass the Anderson-Darling test, while the whitened time-domain samples do not. Overall, glitches cause much larger departures from Gaussianity than adiabatic variation in the noise levels.

To-date, there have been no detailed studies of the effects of non-stationary and non-Gaussian noise on stochastic background analyses beyond the theoretical investigations in [23, 24, 86]. However, a variety of checks have been applied to the LIGO-Virgo analyses using time-shifted data and hardware and software signal injections, and the results were found to be consistent with the performance expected for stationary, Gaussian noise [7, 8]. In particular, the distribution of the residuals of the cross-correlation detection statistic, formed by subtracting the mean and scaling by the square root of the variance, have been shown to be Gaussian distributed [8].

9.3 Gaps and irregular sampling

Data gaps and irregular sampling do not significantly impact the analyses of interferometer data, but pose a major challenge to pulsar timing analyses.

Interferometer data are regularly sampled, and gaps in the data pose no great challenge since the non-stationarity already demands that the analysis be performed on short segments of coincident data. The main difficulty working with short segments of data is accounting for the filters that need to be applied to suppress spectral leakage [7, 108].

The collection of pulsar timing data is constrained by telescope, funding, and personnel availability. A large number of pulsars are now observed fairly regularly, with observations occurring every 2–3 weeks. Older data sets are less regularly sampled, and often have gaps of months or even years [172]. Moreover, the sensitivity of the instruments varies significantly over time, making the data highly non-stationary, thus obviating the benefit of performing the analyses in the frequency domain. For these reasons, modern pulsar timing analyses are conducted directly in the time domain.

Noise modeling for pulsar timing has become increasingly more sophisticated [110, 172], but in broad strokes, the two main terms in the noise model are the measurement errors σ_i in each time-of-arrival measurement, which are assumed to be uncorrelated between time samples i and j, and a stationary red noise component S_{ij} that depends on the lag $|i - j|$ [189]. These contribute to the time-domain noise correlation matrix C_n, which appears in the Gaussian likelihood (3.47):

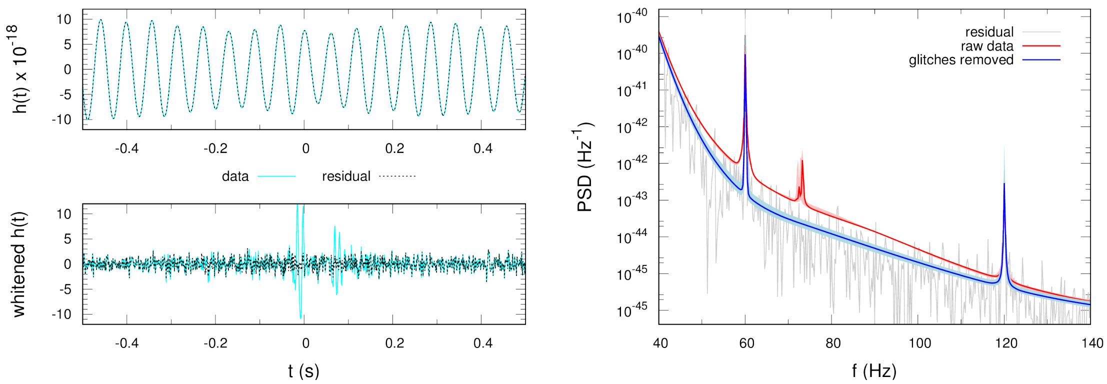
$$(C_n)_{ij} = \sigma_i^2 \delta_{ij} + S_{ij} \,. \tag{9.7}$$

The data gaps and irregular sampling imply that the time lags $|i-j|$ take on a wide range of values, and do not come in multiples of a fixed sample rate Δt. Inverting the large noise matrix C_n to compute the likelihood can be very expensive unless clever tricks are used [190, 191].

9.4 Advanced noise modeling

The traditional approach to noise modeling has been to assume a simple model, such as the noise being stationary and Gaussian, and then measure the consequences this has on the analyses using Monte Carlo studies of time-shifted data and simulated signals. An alternative approach is to develop more flexible noise models that can account for various types of non-stationarity and non-Gaussianity.

One such approach is the *BayesWave/BayesLine* algorithm, which uses a two-part noise model composed of a stationary, Gaussian component $S(f)$, and short duration glitches, $g(t)$, modeled as Gaussian-enveloped sinusoids [48, 113]. The spectral model $S(f)$ is based on a cubic-spline with a variable number of control points to model the smoothly-varying part of the spectrum, and a collection of truncated Lorentzians to model sharp line features. The optimal number and placement of the control points and Lorentizians is determined from the data using a trans-dimensional Markov Chain Monte Carlo technique. The same technique is used to determine the number of sine-Gaussian glitches and their parameters (central time and frequency, duration, etc.). This approach has been applied to both LIGO data [48, 113] and pulsar timing data [61]. Figure 74 demonstrates the application of the *BayesWave* and *BayesLine* algorithms to data from the LIGO Hanford detector during the S6 science run of the initial LIGO detectors. Removing the glitches has a significant impact on the inferred power spectra. Figure 75 displays histograms of the whitened Fourier coefficients for the data shown in Figure 74 with and without glitch removal.



(a) (b)

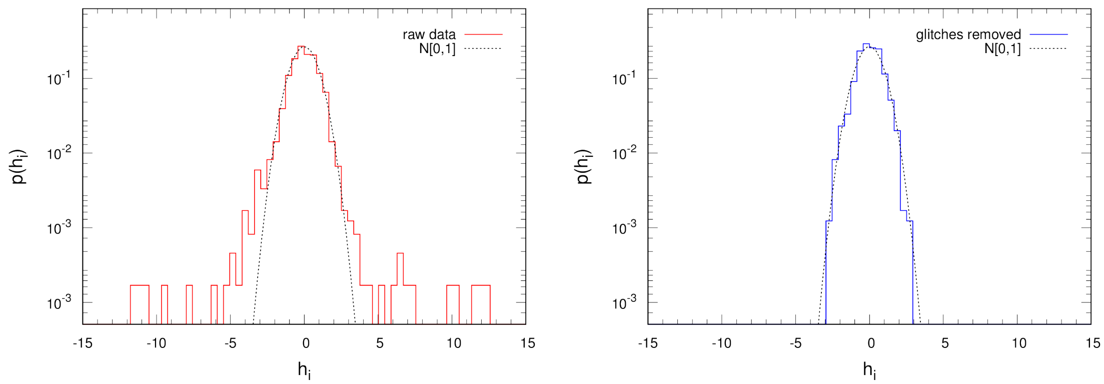
Figure 74: Panel (a) shows a 1-second sample of LIGO S6 data. The upper plot shows the raw data and the lower plot shows the data whitened by the median *BayesLine* spectra with glitch subtraction by the *BayesWave* algorithm. The solid aqua line is the data before glitch removal, and the dotted black line is after glitch removal. Panel (b) shows the median and 90% credible bands for the spectral model with (blue) and without (red) glitch subtraction. The grey line shows the power spectra of the data after glitch removal. Figures provided by Tyson Littenberg.

Additional models for non-stationary and non-Gaussian noise have been considered by several authors. The detection of deterministic and stochastic signals was considered in [23, 24, 86] for a variety of non-Gaussian noise models, including exponential and two-component Gaussian models. The two-component Gaussian model combined with a non-stationary glitch model was studied in [112]. Student's t-distribution was considered in [145]. A non-stationary and non-Gaussian noise model was derived in [138] based on a Poisson distribution of sine-Gaussian glitches.

9.5 Correlated noise

The standard cross-correlation statistic for detecting stochastic backgrounds relies on the noise in each detector being uncorrelated. If we return to the simple model for colocated and coaligned detectors, with white Gaussian noise and a white Gaussian signal (Section 4.3.1), but now introduce a correlated noise component $S_{n_{12}}$, then the correlation matrix for the signal-plus-noise model becomes

$$C = \left[\begin{array}{cc} (S_{n_1} + S_h)\, \mathbb{1}_{N\times N} & (S_h + S_{n_{12}})\, \mathbb{1}_{N\times N} \\ (S_h + S_{n_{12}})\, \mathbb{1}_{N\times N} & (S_{n_2} + S_h)\, \mathbb{1}_{N\times N} \end{array} \right] , \tag{9.8}$$



(a) (b)

Figure 75: Histograms of the whitened time domain data shown in Figure 74. Panel (a) is without glitch subtraction, while panel (b) is with glitch subtraction. Figures provided by Tyson Littenberg.

yielding the maximum likelihood solution

$$\begin{aligned}
\hat{S}_h &\equiv \frac{1}{N}\sum_{i=1}^{N} d_{1i}d_{2i} - S_{n_{12}}\,, \\
\hat{S}_{n_1} &\equiv \frac{1}{N}\sum_{i=1}^{N} d_{1i}^2 - \hat{S}_h\,, \\
\hat{S}_{n_2} &\equiv \frac{1}{N}\sum_{i=1}^{N} d_{2i}^2 - \hat{S}_h\,.
\end{aligned} \tag{9.9}$$

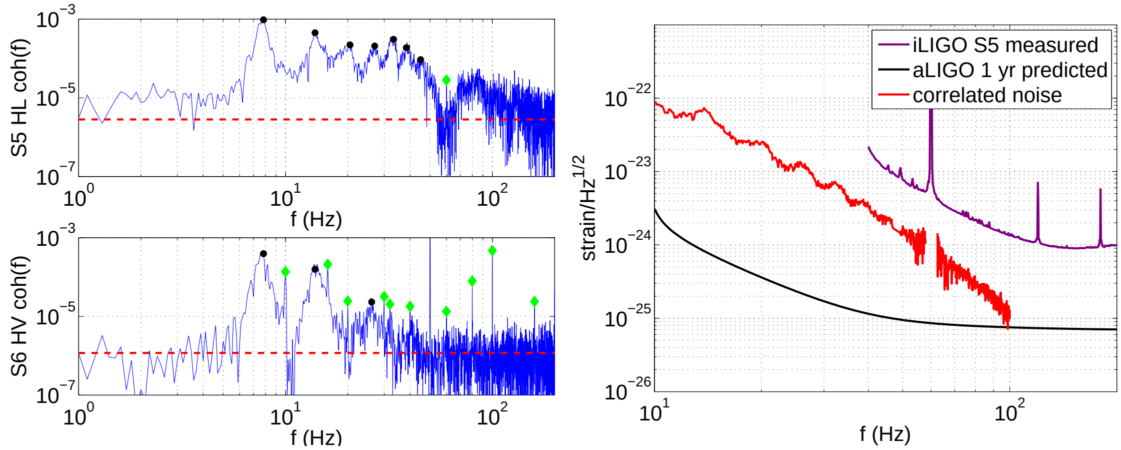
We see that there is a degeneracy between the estimate for the signal $\hat{S}_h$ and the correlated noise $S_{n_{12}}$, with no way to separate the two components. Correlated noise with the same spectrum as the signal presents a *fundamental* limit to the detection of stochastic signals.

If the spectral shape of either, or preferably both, the signal and the correlated noise are known, then it is possible to separate the contributions using techniques similar to those that are used to separate the primordial cosmic microwave background signal from foreground contamination [38]. When the cause of the correlated noise is not fully understood, or when searching for signals with arbitrary spectral shapes, spectrum-based component separation will not be possible.

Several sources of correlated noise have been hypothesized, and in some cases observed, for both interferometer and pulsar timing analyses. Some of the correlations are due to the electronics [7], such as correlations between harmonics of the 60 Hz AC power lines between the LIGO Hanford and LIGO Livingston detectors, and correlations at multiples of 16 Hz from the data sampling referenced to clocks on the Global Positioning System satellites. These narrow-band correlations are easily removed using notch filters. Correlations in the global time standard can also impact pulsar timing observations, as can errors in the ephemeris used in the timing model.

9.5.1 Schumann resonances

One possible broad-band source of correlated noise for ground-based interferometers that has received considerable attention [176, 177, 52] are Schumann resonances in the Earth's magnetic field caused by lightning strikes. These resonances can produce coherent oscillations over thousands of kilometers, and have been observed to produced correlations in magnetometer readings at the LIGO and Virgo sites [176], as shown in panel (a) of Figure 76. The spectrum of the correlations induced in the detector output depend on both the spectrum of the time-varying magnetic field, and the couplings to the instrument. The estimated spectrum of correlated noise in the initial LIGO detectors from Schumann resonances is shown in panel (b) of Figure 76. The estimated spectrum lies below the initial



(a) (b)

Figure 76: Panel (a) shows the cross-correlation of magnetometer readings between the LIGO-Hanford and LIGO-Livingston sites (HL), and also the LIGO-Hanford and Virgo sites (HV). The peaks indicated by black dots are due to Schumann resonances. The green dots mark peaks that are due to correlations caused by the electronics. Panel (b) shows the amplitude spectra of the initial and advanced LIGO detectors compared to the estimated level of the correlated noise due to Schumann resonances. The correlated noise level in advanced LIGO should be lower due to differences in the design (notably the lack of magnets attached to the mirrors). Both figures are taken from [176].

LIGO noise curve, but above the design noise curve for the advanced instruments. The situation is not as dire as it looks, however, since the advanced LIGO detectors have a different design that should have weaker coupling to magnetic fields. Nonetheless, Schumann resonances may end up being a limiting factor for advanced LIGO stochastic searches, and efforts are underway to model and subtract their effects [52]. Correlated noise is a much larger problem for colocated detectors, such as the 2 km and 4 km initial LIGO detectors that shared the Hanford site. There it was found that correlated noise prevented the data at frequencies below 460 Hz from being used for stochastic background searches [4].

Perhaps the greatest challenge comes from correlated noise sources of unknown origin. Such noise sources may be well below the uncorrelated noise level in each detector, and thus very hard to detect outside of the cross-correlation analysis. One way of separating these noise sources from a stochastic signal is to build a large number of interferometers at many locations around the world. Each pair of detectors will then have a unique overlap function for gravitational-wave signals that will differ from the spatial correlation pattern of the noise (unless we are incredibly unlucky!). In principle, the difference in the frequency-dependent spatial correlation patterns of the signal and the noise will allow the two components to be separated.

9.6 What can one do with a single detector (e.g., eLISA)?

The discovery of the cosmic microwave background was described in a paper with the unassuming title "A Measurement of Excess Antenna Temperature at 4080 Mc/s" [133]. Penzias and Wilson used a single microwave horn, and announced the result after convincing themselves that no instrumental noise sources, including pigeon droppings, could be responsible for the excess noise seen in the data. In principle, the same approach could be used to detect a stochastic gravitational-wave signal using a single instrument.

Single-detector detection techniques will be put to the test when the first space-based gravitational-wave interferometer is launched, since (unless the funding landscape changes dramatically) the instrument will be a single array of 3 spacecraft. Assuming that pairs of laser links operate between each pair of spacecraft, it will be possible to synthesize multiple interferometry signals from the phase readouts [62]. One particular combination of the phase readouts, called the T channel, corresponds to a Sagnac interferometer, and is relatively insensitive to low-frequency gravitational waves, forming an approximate null channel (see Section 4.7 for a discussion of null channels). Other combinations, such as the so-called A and E channels [137], are much more sensitive to gravitational-wave signals. Using the Sagnac T to measure the instrument noise, the relative power levels in the $\{A, E, T\}$ channels can be used to separate a stochastic signal from instrument noise [180].

LISA-type observatories operate as synthetic interferometers by forming gravitational-wave observables in post-processing using different combinations of the phasemeter read-outs from each inter-spacecraft laser link. The combinations synthesize effective equal-path-length interferometers to cancel the otherwise overwhelming laser frequency noise. These combinations have to account for the unequal and time-varying distances between the spacecraft.

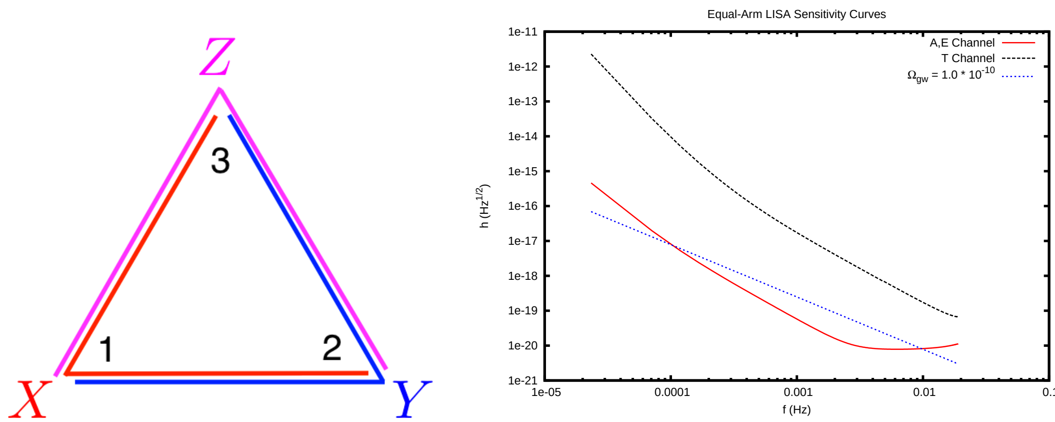
In the conceptually simpler equal-arm-length limit, the Michelson-type signal extracted from vertex 1 (see panel (a) of Figure 77) is given by

$$X(t) = M_1(t) - M_1(t - 2L)\,, \tag{9.10}$$

where

$$M_1(t) = \Phi_{12}(t - L) + \Phi_{21}(t) - \Phi_{13}(t - L) - \Phi_{31}(t)\,, \tag{9.11}$$

and $\Phi_{ij}(t)$ is the readout from the phasemeter on spacecraft j that receives light from



(a) (b)

Figure 77: Panel (a) shows the geometry of a LISA-like space interferometer and the laser paths for the synthetic Michelson interferometers X, Y, Z. Panel (b) shows sensitivity curves for the A, E, T interferometry variables compared to a scale-invariant background, $\Omega_{\rm gw}(f) = \Omega_0 = {\rm const}$, with $\Omega_0 = 10^{-10}$. The Sagnac-like T channel is far less sensitive than the Michelson-like A, E channels, and can be used to measure the instrument noise levels. (Panel (b) is adapted from [16]).

spacecraft i. Permuting the spacecraft labels $\{1, 2, 3\}$ yields equivalent expressions for the Michelson observables Y and Z, as shown in panel (a) of Figure 77. The phasemeter readouts $\Phi_{ij}(t)$ are impacted by acceleration noise S^a_{ij} and position noise S^p_{ij}. When the noise levels in each spacecraft are equal, there exist noise-orthogonal combinations [137]:

$$\begin{aligned} A &\equiv \frac{1}{3}(2X - Y - Z)\,, \\ E &\equiv \frac{1}{\sqrt{3}}(Z - Y)\,, \\ T &\equiv \frac{1}{3}(X + Y + Z)\,. \end{aligned} \tag{9.12}$$

Note that these variables are only noise-orthogonal in the symmetric noise limit. For example, the position noise contribution to the cross-spectra $\langle AE \rangle$ is given by

$$\langle AE \rangle = -\frac{4}{3\sqrt{3}} \sin^2\left(\frac{f}{f_*}\right)\left(2\cos\left(\frac{f}{f_*}\right) + 1\right)\left(S^p_{13} - S^p_{12} + S^p_{31} - S^p_{21}\right), \tag{9.13}$$

which vanishes when $\{S^p_{13}, S^p_{12}, S^p_{31}, S^p_{21}\}$ are equal, but not otherwise [16]. The synthetic interferometers A, E are rotated by 45 degrees with respect to each other, and provide instantaneous measurements of the $+$ and $\times$ polarization states. The Sagnac-like T channel is relatively insensitive to gravitational waves for frequencies below the transfer frequency

$f_* \equiv c/(2\pi L)$. The T channel can be used to infer the instrument noise level, so that any excess in the A, E channels can then be confidently attributed to gravitational waves [180]. For frequencies $f \ll f_*$ the $\{A, E, T\}$ channels have uncorrelated responses to unpolarized, isotropic stochastic gravitational-wave signals.

There are some subtleties associated with using the T channel as a noise reference as the noise combinations in T differ from those in A, E. For example, the acceleration noise appears in T as [16]:

$$\langle TT \rangle = \frac{16}{9} \sin^2\left(\frac{f}{f_*}\right)\left(1 - \cos\left(\frac{f}{f_*}\right)\right)^2 \left(S^a_{12} + S^a_{13} + S^a_{31} + S^a_{32} + S^a_{23} + S^a_{21}\right), \quad (9.14)$$

while the acceleration noise appears in A and E as

$$\begin{aligned} \langle AA \rangle \;\; = \;\; & \frac{16}{9} \sin^2\left(\frac{f}{f_*}\right)\Bigg\{ \cos\left(\frac{f}{f_*}\right)\left[4\big(S^a_{12} + S^a_{13} + S^a_{31} + S^a_{21}\big) - 2\big(S^a_{23} + S^a_{32}\big)\right] \\ & + \cos\left(\frac{2f}{f_*}\right)\left[\frac{3}{2}(S^a_{12} + S^a_{13} + S^a_{23} + S^a_{32}) + 2(S^a_{31} + S^a_{21})\right] \\ & + \frac{9}{2}(S^a_{12} + S^a_{13}) + 3(S^a_{31} + S^a_{21}) + \frac{3}{2}(S^a_{23} + S^a_{32}) \Bigg\}, \end{aligned} \quad (9.15)$$

and

$$\begin{aligned} \langle EE \rangle = \frac{16}{3} \sin^2\left(\frac{f}{f_*}\right) \Bigg\{ & S^a_{23} + S^a_{32} + S^a_{21} + S^a_{31} + 2\cos\left(\frac{f}{f_*}\right)\left(S^a_{23} + S^a_{32}\right) \\ & + \cos^2\left(\frac{f}{f_*}\right)\left(S^a_{23} + S^a_{32} + S^a_{12} + S^a_{13}\right)\Bigg\}. \end{aligned} \quad (9.16)$$

In the ideal case where the noise levels are the same in each link, T provides a measurement of the average noise, which can then be used as an estimator for the noise in A, E. An analysis that assumes common noise levels will overstate the sensitivity to a signal. A more conservative approach is allow for unequal noise levels and to infer the individual contributions from the data. For example, if one link is particularly noisy, it will dominate the noise in T, and enter unequally in A and E, making it possible to identify the bad link and account for it in the analysis.

Bayesian inference is ideally suited to the task of jointly inferring the signal and noise levels using models that fold in prior knowledge of the signals and instrument components [16]. The separation is aided by the difference in the transfer functions for the signal and the noise. Analytic expressions for the signal transfer or auto-correlated response functions (see Section 5.3) can be derived in the low-frequency limit $f \ll f_*$:

$$\mathcal{R}_{TT} = 4\sin^2\left(\frac{f}{f_*}\right)\left[\frac{1}{12096}\left(\frac{f}{f_*}\right)^6 - \frac{61}{4354560}\left(\frac{f}{f_*}\right)^8 + ...\right], \quad (9.17)$$

and

$$
\mathcal{R}_{AA} = \mathcal{R}_{EE} = 4\sin^2\left(\frac{f}{f_*}\right)\left[\frac{3}{10} - \frac{169}{1680}\left(\frac{f}{f_*}\right)^2 + \frac{85}{6048}\left(\frac{f}{f_*}\right)^4 - \frac{178273}{159667200}\left(\frac{f}{f_*}\right)^6 + \frac{19121}{24766560000}\left(\frac{f}{f_*}\right)^8 + \ldots\right]. \quad (9.18)
$$

Note that these signal transfer functions are very different from the acceleration noise transfer functions given in (9.14), (9.15), (9.16). The difference in the transfer functions, combined with priors on the functional form of the power spectral density of the noise and signal, allows for the detection of signals that are significantly below the noise level, even when there are not enough links to form the T channel [16]. The sensitivity decreases for less informative priors. In the limit that the priors allow for arbitrarily complicated functional forms for the noise and signal spectra—forms so *contrived* that they can compensate for the differences in the transfer functions—it becomes impossible to separate signal from noise. In practice, a combination of pre-flight and on-board testing, combined with physical modeling, will hopefully constrain the noise model sufficiently to inform the analysis and allow for component separation.

An additional complication for space interferometers operating in the mHz frequency range are the millions of astrophysical signals that can drown-out a cosmologically-generated stochastic background. While the brightest signals from massive black hole mergers, stellar captures, and galactic binaries can be identified and subtracted, a large number of weaker overlapping signals will remain, creating a residual *confusion noise*. The largest source of confusion noise is expected to come from millions of compact white-dwarf binaries in our galaxy. The annual modulation of the white-dwarf confusion noise due to the motion of the LISA spacecraft (see Figure 39) will allow for this component to be separated from an isotropic stochastic background, though at the cost of reduced sensitivity to the background [17].

10 Prospects for detection

It's tough to make predictions, especially about the future. *Yogi Berra*

The detection of the binary black hole merger signals GW150914 and GW151226 give us confidence that stochastic gravitational waves will be detected in the not-to-distant future. Not only do they show that our basic measurement principles are sound, they also point to the existence of a much larger population of weaker signals from more distant sources that will combine to form a stochastic background that may be detected by 2020 [10]. Indeed, a confusion background from the superposition of weaker signals eventually becomes the limiting noise source for detecting individual systems [36]. As a general rule of thumb, individual bright systems will be detected before the background for transient signals (those that are in-band for a fraction of the observation time), while the reverse is true for long-lived signals, such as the slowly evolving supermassive black-hole binaries targeted by pulsar timing arrays [144]. The prospects for detecting more exotic stochastic signals, such as those from phase transitions in the early Universe or inflation, are much less certain, but are worth pursing for their high scientific value. In this section we begin with a brief review of detection sensitivities curves across the gravitational-wave spectrum, followed by a review of the current limits and prospects for detection in each observational window.

10.1 Detection sensitivity curves

Detector sensitivity curves provide a useful visual indicator of the sensitivity of an instrument to potential gravitational-wave sources. A good pedagogical description of the various types of sensitivity curve in common use can be found in [123]. Here we provide a more condensed summary.

The simplest type of sensitivity curve is a plot of the power spectral density of the detector noise $P_n(f)$, or its amplitude spectral density $\sqrt{P_n(f)}$. But these plots can be misleading since they do not take into account the frequency-dependent response to gravitational waves seen in Figure 32. A better quantity to plot is the sky and polarization-averaged amplitude spectral density

$$h_{\rm eff}(f) \equiv \sqrt{S_n(f)} = \sqrt{P_n(f)/\mathcal{R}(f)}\,, \tag{10.1}$$

which has units of strain/$\sqrt{\rm Hz}$, or the corresponding (dimensionless) characteristic strain noise

$$h_n(f) \equiv \sqrt{f S_n(f)}\,, \tag{10.2}$$

where $\mathcal{R}(f) \equiv \Gamma_{II}(f)$ is the transfer function defined in (5.44) and (5.45). Figure 78 shows the construction of a LISA sensitivity curve from $P_n(f)$ and $\mathcal{R}(u)$, where $u = f/f_*$ and $f_* = c/(2\pi L)$. Note that for LIGO the factor $\mathcal{R}(f)$ is usually not included in sensitivity plots since $f_* \simeq 12\,{\rm kHz}$, and $\mathcal{R}(f)$ is effectively constant across the LIGO band.

The amplitude spectral density sensitivity curve $h_{\rm eff}(f)$ has to be interpreted with some care, as simply comparing this curve to the amplitude spectral density of a signal does not immediately convey how detectable the signal is, as the likelihood and detection statistics derived from the likelihood involve integrals over frequency. The problem is compounded

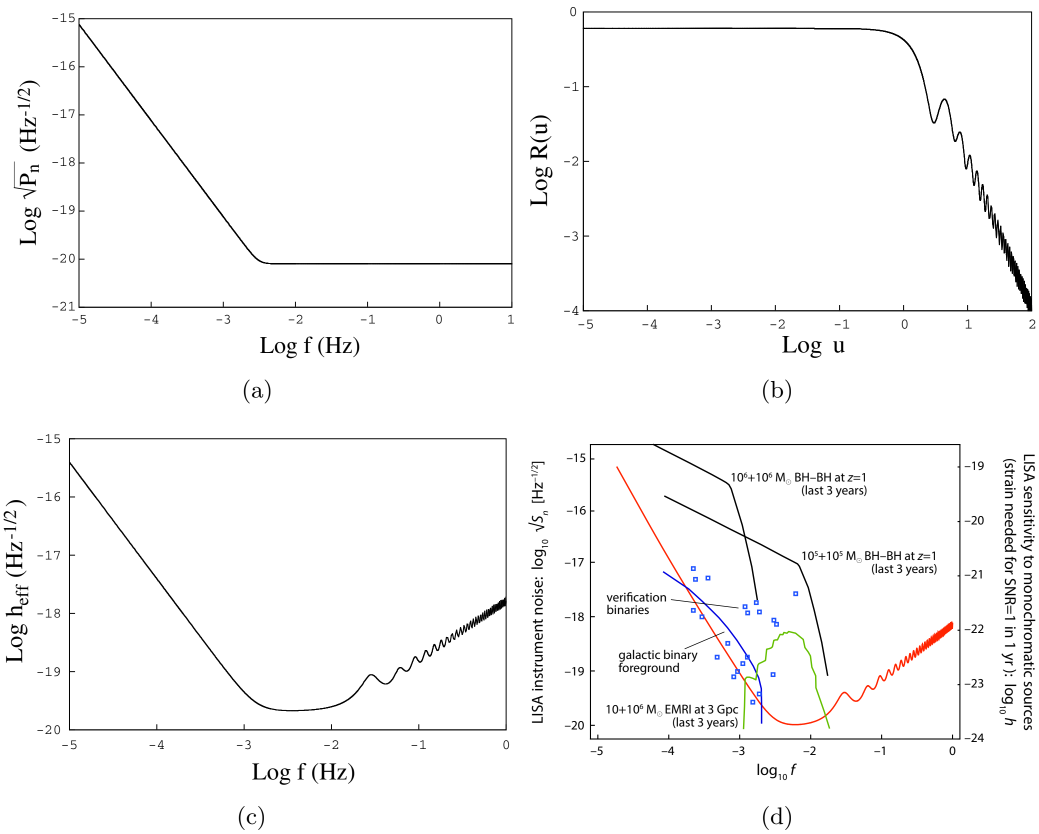


Figure 78: Constructing a sensitivity curve for the LISA detector. Panel (a) shows the amplitude spectral density of the noise. Panel (b) shows the sky and polarization averaged response function. Panel (c) shows the sensitive curve found by dividing the noise spectral density by the response function. Panel (d) compares the filtered effective signal strength $\sqrt{2fTS_h(f)}$ for various signals to the LISA sensitivity curve $\sqrt{S_n(f)}$. The figures in panels (a)–(c) are taken from [106]. The figure in panel (d) was provided by M. Vallisneri.

by the necessity to plot the sensitivity curves on a log-log scale, where "integration-by-eye" misses the increase in the number of frequency bins per logarithmic frequency interval. Rather than plot the raw signals, it is more informative to show quantities that account for the detection techniques being used. For example, the signal-to-noise ratio ρ for a deterministic signal $\tilde{h}(f)$ is given by

$$\rho^2 = \int_0^\infty \frac{4|\tilde{h}(f)|^2}{P_n(f)}\, df = \int_{-\infty}^\infty \frac{4f|\tilde{h}(f)|^2}{P_n(f)}\, d\ln f \,. \tag{10.3}$$

Averaging over sky location and polarization we have

$$\overline{\rho^2} = \int_{-\infty}^\infty \frac{4f\tilde{h}^2_{\rm rss}(f)\mathcal{R}(f)}{P_n(f)}\, d\ln f = \int_{-\infty}^\infty \frac{(2fT)S_h(f)}{S_n(f)}\, d\ln f \,, \tag{10.4}$$

where $\tilde{h}^2_{\rm rss}(f) \equiv |\tilde{h}_+(f)|^2 + |\tilde{h}_\times(f)|^2$, and $S_h(f)$ is the power spectral density of the gravitational-wave signal,

$$S_h(f) \equiv \frac{2\tilde{h}^2_{\rm rss}(f)}{T} \,. \tag{10.5}$$

The quantity $(2fT)S_h(f)/S_n(f)$ is the contribution to the square of the signal-to-noise ratio per logarithmic frequency interval. The factor of $2fT$ describes the boost that we get by coherently integrating the signal over many cycles. For deterministic signals the signal-to-noise ratio grows as $T^{1/2}$. Since sensitivity curves are usually plotted in terms of the amplitude spectral density $h_{\rm eff}(f) = \sqrt{S_n(f)}$, it is natural to plot signals in terms of the square-root of the numerator of (10.4). Representative LISA sources are represented in this way in panel (d) of Figure 78. An alternative choice is to plot both of these quantities multiplied by the square-root of the frequency, which yield the characteristic strain for the signal, $h_c(f)$, as well as for the noise, $h_n(f)$. Examples of characteristic strain sensitivity curves are shown in Figure 79.

For isotropic stochastic signals, the sky location and polarization-averaged signal-to-noise ratio ρ is

$$\rho^4 = 2T\int_0^\infty df \sum_{I=1}^M \sum_{J>I}^M \frac{\Gamma^2_{IJ}(f)S^2_h(f)}{P_{n_I}(f)P_{n_J}(f)} = \int_{-\infty}^\infty \frac{(2fT)S^2_h(f)}{S^2_{\rm net}(f)}\, d\ln f \,,$$

where

$$S_{\rm net}(f) \equiv \left[\sum_{I=1}^M \sum_{J>I}^M \frac{\Gamma^2_{IJ}(f)}{P_{n_I}(f)P_{n_J}(f)}\right]^{-1/2} \,. \tag{10.6}$$

For stochastic signals, the signal-to-noise ratio grows as $T^{1/4}$ (this assumes that we are in the weak-signal limit, and that the effective low-frequency cutoff does not change with time. See [159] for a more complicated scaling that occurs for pulsar timing arrays.) Following the same logic as was applied to deterministic signals, it would be natural to plot $(2fT)^{1/4}\sqrt{S_h(f)}$ against sensitivity curves defined by $\sqrt{S_{\rm net}(f)}$. Unfortunately, such conventions are not uniformly applied, and the factor of $(2fT)^{1/4}$ is often applied to $\sqrt{S_{\rm net}(f)}$ instead:

$$h_{\rm eff}(f) \equiv \frac{1}{(2Tf)^{1/4}}\sqrt{S_{\rm net}(f)} \,. \tag{10.7}$$

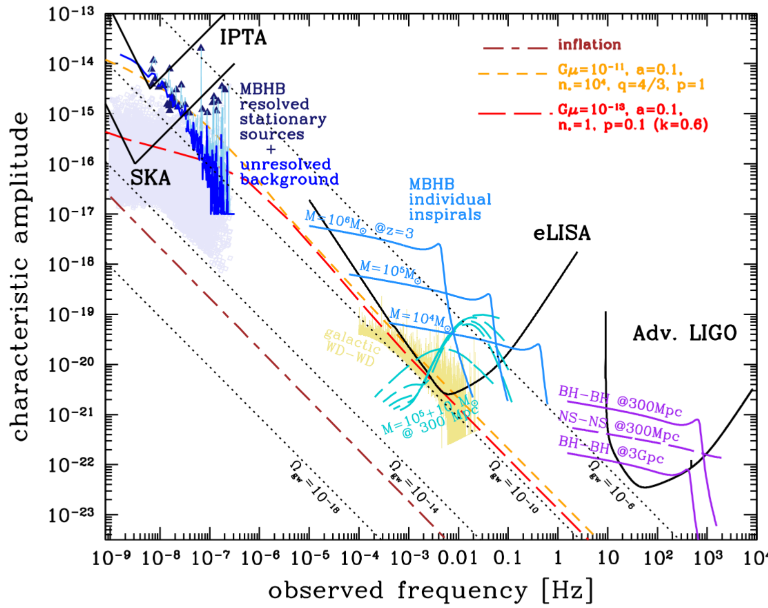


Figure 79: Examples of detector sensitivity curves compared to potential gravitational-wave signals, comparing the characteristic strain signal $h_c(f)$ to the characteristic strain noise $h_n(f)$. Figure taken from [95].

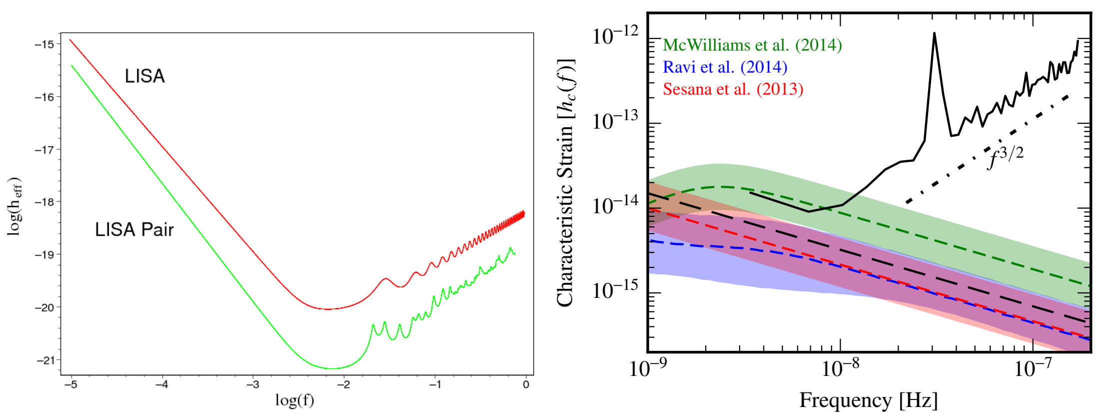
A plot of $h_{\rm eff}(f)$, averaged over a logarithmic frequency interval $\Delta f = f/10$ for a crossed pair of LISA-like detectors is shown in Figure 80. Also shown in this figure are the related per-frequency-bin upper bounds that are quoted by pulsar timing groups using fixed frequency intervals $\Delta f = 1/T$.

The most common form of sensitivity curve for stochastic backgrounds compares predictions of the gravitational-wave energy density $\Omega_{\rm gw}(f)$ to the equivalent noise energy density $\Omega_n(f) \equiv 2\pi^2 f^3 S_n(f)/(3H_0^2)$. These plots have the advantage of being easy to produce and explain, but they do not fully capture the boost that comes from integrating over frequencies. An alternative form of sensitivity curve that better represents the analysis procedure uses the envelope of limits that can be placed on power-law stochastic backgrounds [178, 124]. This method has the advantage of incorporating the integrated nature of the detection statistic. Examples for advanced LIGO and PTAs are shown in Figure 81.

10.2 Current observational results

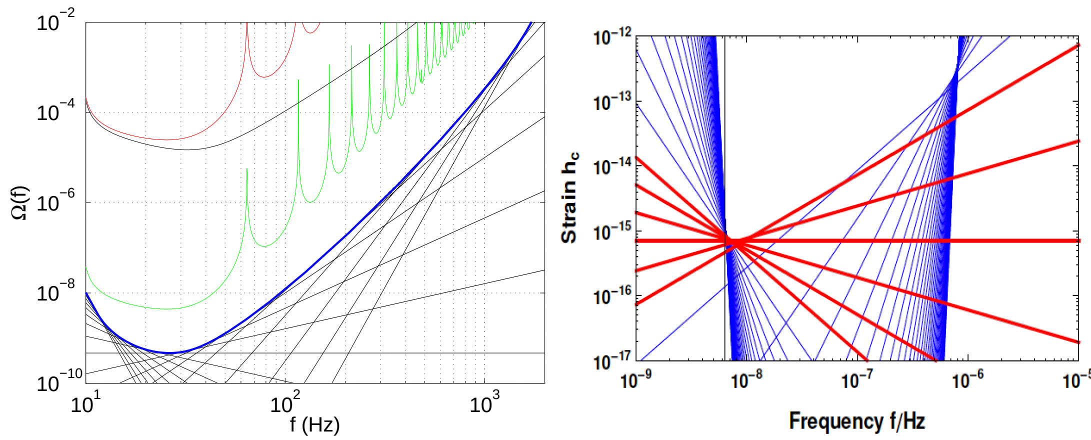
10.2.1 CMB isotropy

The cosmic microwave background (CMB) provides a snapshot of the Universe $\approx 400,000$ years after the big bang. During this epoch, the dense, hot plasma that filled the early Universe dilutes and cools to the point where electrons and ions combine to form a neutral



(a) (b)

Figure 80: Panel (a) compares $h_{\rm eff}(f)$ for an isotropic stochastic background for a single LISA detector to that for a pair of LISA detectors arranged in a crossed-star configuration using an observation time of one year. (Figure taken from [46].) Panel (b) compares the per-frequency-bin ($\Delta f = 1/T$) upper limits on an isotropic stochastic background derived from the NANOGrav 9-year data set (solid black line) to three astrophysical models for the signal from supermassive black hole binaries. The upturn in the bound at low frequencies and the spike at $f = 1/\text{year}$ are due to the timing model acting as a filter on the signal. (Figure taken from [32].)



(a) (b)

Figure 81: Panel (a) shows the sensitivity of the advanced LIGO Hanford-Livingston detector pair in terms of gravitational-wave energy density $\Omega_{\rm gw}(f)$ using a variety of methods. The blue line is the sensitivity to isotropic stochastic signals with power-law spectra, formed from the envelope of backgrounds with a wide range of spectral slopes (shown as straight black lines). Also shown as a black curve is the noise spectral density of a single LIGO detector converted to units of $\Omega(f)$. The red and green lines are variants of $h_{\rm eff}(f)$, again converted to units of $\Omega(f)$. The lower green curve is for an observation time of one year and $\Delta f = 0.25\,{\rm Hz}$. (Figure taken from [178].) Panel (b) shows the characteristic strain sensitivity for a hypothetical pulsar timing array formed from the envelope of a large number of power law models. The red lines show a subset of the power law models used. The upper and lower frequency limits to the sensitivity are set by the observation cadence and the observation time, respectively. (Figure taken from [124].)

gas that is transparent to photons. Maps of the CMB contain a record of the conditions when the CMB photons were last scattered.

Gravitational waves propagating through the early Universe, referred to as tensor perturbations in the CMB literature, can leave an imprint in the temperature and polarization pattern when CMB photons scatter off the tidally-squeezed plasma. The challenge is to separate out the contributions from primordial scalar, vector, and tensor perturbations, and to separate these primordial contributions from subsequent scattering by dust grains and hot gas.

Observations by the *COBE, WMAP* and *Planck* mission, along with a host of ground-based and ballon-borne experiments, have provided strong evidence in support for the inflation paradigm, where the Universe undergoes a short period of extremely rapid expansion driven by some, as yet unknown, *inflaton* field. To keep the discussion brief, we focus our review on the standard single-field "slow-roll" inflation model, and direct the reader to more extensive CMB-focused reviews, e.g., [99], that cover more exotic models.

The rapid expansion of some small patch of the very early Universe will erase any initial anisotropy and inhomogeneity, allowing the patch to be modeled by a flat Friedmann-Lemaitre-Robertson-Walker (FLRW metric) with scale factor $a(t)$. The Einstein equations for a FLRW Universe containing an inflaton field ϕ with potential $V(\phi)$ are given by[22]

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{1}{3M_{\rm Pl}^2}\left(\frac{1}{2}\dot{\phi}^2 + V(\phi)\right) , \tag{10.8}$$

and

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} = 0. \tag{10.9}$$

In the slow-roll regime, the kinetic energy of the inflaton field $\frac{1}{2}\dot{\phi}^2$ is assumed to be much smaller than the potential energy $V(\phi)$, with ϕ having reached "terminal velocity", such that $\ddot{\phi} \ll H\dot{\phi}$. Thus,

$$3H\dot{\phi} \simeq -V_{,\phi} \quad \text{and} \quad H^2 \simeq \frac{V}{3M_{\rm Pl}^2} . \tag{10.10}$$

Necessary conditions for these approximations to hold can be expressed in terms of a Taylor series expansion of the inflaton potential leading to conditions on the first and second derivatives of the potential:

$$\epsilon_V \equiv \frac{M_{\rm Pl}^2 V_{,\phi}^2}{2V^2} \ll 1 , \qquad \eta_V \equiv \frac{M_{\rm Pl}^2 V_{,\phi\phi}}{V} \ll 1 . \tag{10.11}$$

The solution of the Einstein equations for slow-roll inflation is well-approximated by an exponentially de Sitter Universe. Quantum fluctuations in the otherwise smooth inflaton field and gravitational field give rise to scalar and tensor perturbations, which leave their

[22]For our discussion of inflation, we will work in *particle physics units* where both $c = 1$ and $\hbar = 1$. In place of using Newton's gravitational constant G, we will use the *reduced* Planck mass $M_{\rm Pl} \equiv (\hbar c/8\pi G)^{1/2} = 2.435 \times 10^{18}$ GeV$/c^2$. In these units $M_{\rm Pl}^2 = 1/8\pi G$, which simplifies several of the formulae. If you want to reinstate all of the relevant factors of $\hbar$ and c, note that the inflaton field ϕ has dimensions of energy and the inflaton potential $V(\phi)$ has dimensions of energy density.

imprint in the CMB. On large scales the power spectra for the scalar and tensor fluctuations can be written as

$$P_s(k) = A_s \left(\frac{k}{k_*}\right)^{n_s(k)-1} \quad \text{and} \quad P_t(k) = A_t \left(\frac{k}{k_*}\right)^{n_t(k)} \tag{10.12}$$

where the reference wavenumber $k_* = 2\pi/\lambda_*$ is typically chosen to correspond to wavelengths $\lambda_* \sim 100\,\text{Mpc}$. The spectral indices $n_s(k)$ and $n_t(k)$ are usually written in terms of a power series expansion in $\ln k$:

$$n_s(k) = n_s + \frac{1}{2}\frac{dn_s}{d\ln k}\ln\left(\frac{k}{k_*}\right) + \frac{1}{6}\frac{d^2 n_s}{d\ln k^2}\ln\left(\frac{k}{k_*}\right)^2 + \ldots \tag{10.13}$$

The amplitude and spectral indices are related to the energy scale for inflation, V, and the slow-roll parameters ϵ_V and η_V:

$$A_s \simeq \frac{V}{24\pi^2 M_{\text{Pl}}^4 \epsilon_V} \quad \text{and} \quad A_t \simeq \frac{2V}{3\pi^2 M_{\text{Pl}}^4}\,, \tag{10.14}$$

and

$$n_s \simeq 1 + 2\eta_V - 6\epsilon_V \quad \text{and} \quad n_t \simeq -2\epsilon_V\,. \tag{10.15}$$

Measuring A_s, A_t, and n_s fixes the energy scale of inflation, V, and the two leading terms in the Taylor series expansion of the inflaton potential $V_{,\phi}$ and $V_{,\phi\phi}$. Additionally measuring n_t would provide a consistency check for the slow-roll model.

One challenge in measuring $P_s(k)$ and $P_t(k)$ is that the scalar and tensor perturbations both source temperature and polarization anisotropies in the CMB radiation. Another challenge is that foreground gas and dust can also contribute to the temperature and polarization anisotropies. The various components can be teased apart by observing a wide range of CMB energies across a wide range of angular scales.

The primordial contribution to the CMB follows a black-body spectrum, while the dominant foreground contribution from gas and dust have very different spectra. By observing at multiple CMB wavelengths the primordial and foreground contributions can be separated. Separating the scalar and tensor contributions to the primordial contribution to the temperature anisotropies can be achieved by making maps that cover a wide range of angular scales, while separating their contributions to the polarization anisotropies can be achieved by decomposing the signal into curl-free E-modes and divergence-free B-modes, and using measurements made on a wide range of angular scales. For a more in-depth description, see Chapter 27 of the Review of Particle Physics [131].

The scalar and tensor contributions to the large-scale temperature anisotropy can be computed using linear perturbation theory. The anisotropy due to tensor fluctuations arises solely from the gravitational potential differences on the last-scattering surface, while the anisotropy due to scalar fluctuations is more complicated, and include contributions from the excitation of sound waves in addition to variations in the gravitational potential. As the co-moving horizon grows, tensor modes that have wavelengths shorter than the horizon size redshift and lose energy. Consequently, the tensor contribution to the CMB anisotropy drops by roughly two orders of magnitude between angular scales $\ell = 2$ and

$\ell = 200$, while the scalar contribution, after an initial dip, grows until reaching the first acoustic peak at $\ell \simeq 220$. Plots of the predicted scalar and tensor contributions to the tenperature (TT) power spectra using the best fit ΛCDM model from *Planck* are shown in panel (a) of Figure 82. By comparing the CMB anisotropy at very large scales ($\ell \sim$ 2–10) and degree scales ($\ell \sim 200$), it is possible to constrain the *tensor-to-scalar ratio* [102]:

$$r \equiv \frac{A_t}{A_s}\,. \tag{10.16}$$

In practice, a more sophisticated joint analysis is performed using all available CMB data (often combined with other data sets, such as maps of large-scale structure, weak lensing, and measurements of the expansion history), simultaneously fitting for a large number of cosmological parameters. The *Planck* temperature map, combined with weak lensing data, provide a precise measurement for the amplitude and spectral index of the scalar perturbations:

$$\ln A_s = -19.928 \pm 0.057\,, \qquad n_s = 0.9603 \pm 0.0073\,, \tag{10.17}$$

and a bound on the tensor-to-scalar ratio:

$$r < 0.12 \quad (95\% \text{ confidence})\,, \tag{10.18}$$

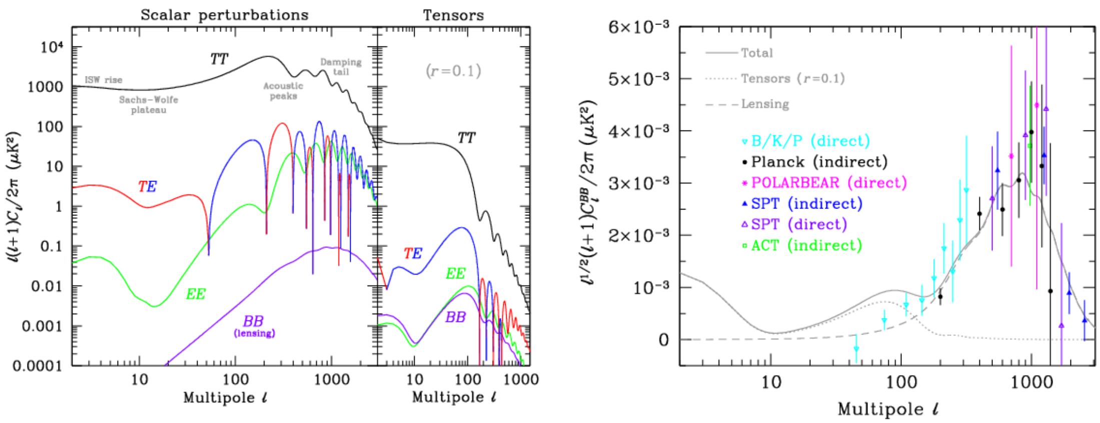
using a pivot scale of $k_* = 0.002\,\mathrm{Mpc}^{-1}$. The Planck bound on r is the most stringent possible using CMB temperature data [102]. (In fact, it beats the theoretical limit slightly since the analysis also used weak lensing and *WMAP* polarization data.) In order to improve on this bound, or to detect the tensor contribution, CMB polarization data must also be used.

The *Planck* bound on r can be mapped into constraints on the gravitational-wave energy density via [182, 107]:

$$\Omega_{\rm gw}(f) = \frac{3rA_s\Omega_r}{128}\left(\frac{f}{f_*}\right)^{n_t}\left[\frac{1}{2}\left(\frac{f_{\rm eq}}{f}\right)^2 + \frac{16}{9}\right]\,, \tag{10.19}$$

where $f = ck/(2\pi)$, $f_{\rm eq} \equiv \sqrt{2}H_0\Omega_m/(2\pi\sqrt{\Omega_r})$ is the frequency of a horizon-scale mode when matter and radiation have the same density, and Ω_m and Ω_r are the matter and radiation density today, in units of the critical density. The projected *Planck* bound from the B-mode power spectrum, along with existing and projected bounds from pulsar timing and aLIGO are shown in Figure 83, which is taken from [107]. Also shown are curves for theoretical models with a large tensor-to-scalar ratio ($r = 0.11$) and a range of spectral tilts n_t.

Coherent motion in the primordial plasma can polarize the CMB photons through Thomson scattering. Scalar perturbations source curl-free E-mode polarization anisotropies, while the tensor modes source divergence-free B-mode polarization anisotropies, in addition to E-modes. In principle, by decomposing the polarization into E and B components, and using observations across a range of angular scales, it should be possible to separate the scalar and tensor contributions. In practice, the measurements are extremely challenging due to the weakness of the signals (nano-Kelvin or smaller polarization fluctuations



(a) (b)

Figure 82: Panel (a) shows the theoretical predictions for the temperature and polarization cross-spectra from scalar and tensor perturbations for the best fit ΛCDM model from *Planck* assuming a tensor-to-scalar ratio of $r = 0.1$. The curves are labeled by type: TT labels the temperature power spectrum, while TE labels the temperature-E-mode cross spectrum and so on. Panel (b) compares recent measurements of the BB spectrum to the theoretical prediction. Both panels are taken from the Review of Particle Physics [131].

as compared to micro-Kelvin temperature fluctuations) and foreground noise. The main noise contributions come from gravitational lensing, which converts a fraction of the much larger E-mode anisotropy into B-modes, and scattering by dust grains, which can convert unpolarized CMB radiation into E and B modes. Both of these potential noise sources have recently been detected [79, 20]. The detection of B-mode polarization on large angular scales by *BICEP2* was originally interpreted as having a primordial original [19], but a joint analysis using *Planck* dust maps [20] showed the signal to be consistent with foreground noise.

While detecting the primordial B-mode contribution is very challenging, the pay-off is very large, as measuring the amplitude of the tensor perturbations, A_t, fixes the energy scale of inflation, and can be used to strongly constrain models of inflation.

10.2.2 Pulsar timing

Pulsar timing observations have made tremendous progress in the past ten years and are now producing limits that seriously constrain astrophysical models for supermassive black hole mergers. The current observations are most sensitive at $f \sim 10^{-8}$ Hz, so we choose a reference frequency of $f_{\rm ref} = 10^{-8}$ Hz, and quote the latest bounds on $\Omega_{\rm gw}(f) = \Omega_\beta (f/f_{\rm ref})^\beta$ in terms of bounds on Ω_β for a Hubble constant value of $H_0 = 70\ {\rm km\,s^{-1}\,Mpc^{-1}}$.

For a scale invariant ($n_t = 0$) cosmological background, $\beta = 0$. The most recent 95%

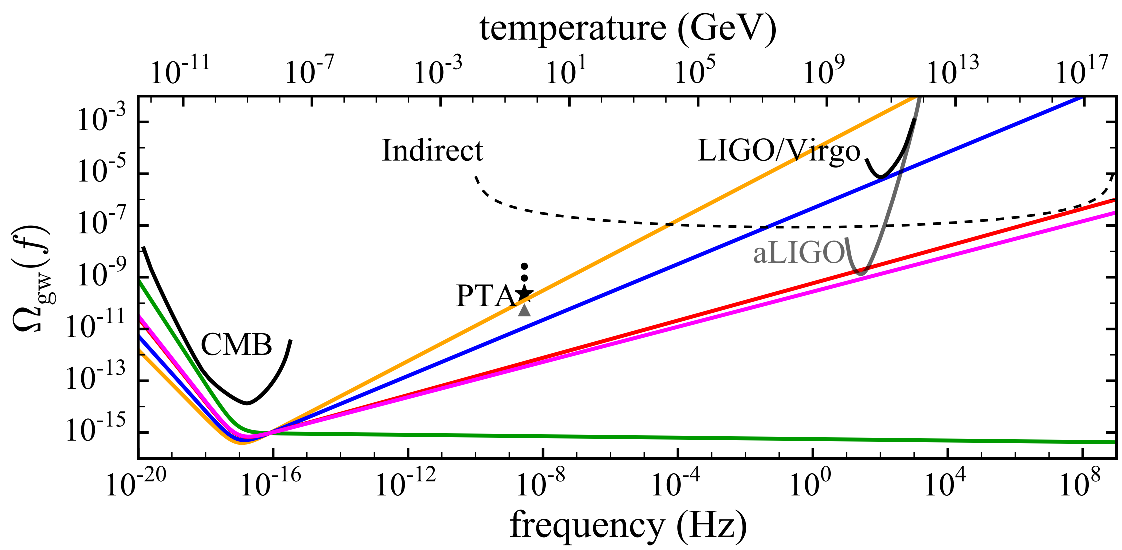


Figure 83: Current and projected bounds on $\Omega_{\rm gw}(f)$ from CMB measurements, pulsar timing observations and ground based interferometers. The curve marked "CMB" shows the projected sensitivity of the *Planck* satellite to primordial B-mode polarization anisotropies. The black star marked "PTA" is the current 95% upper limit from the Parkes pulsar timing array. The LIGO and aLIGO sensitivity curves were produced using the power-law envelope method [178]. The curved labeled "indirect bounds" was produced by converting bounds on the total gravitational-wave energy density from CMB temperature and polarization power spectra, weak lensing, baryon acoustic oscillations, and Big Bang nucleosynthesis to bounds on the energy density per logarithmic frequency interval using power law models. The colored lines are theoretical predictions for the primordial background assuming $r = 0.11$ and for spectral slopes $n_t = 0.68$ (orange curve), $n_t = 0.54$ (blue), $n_t = 0.36$ (red), and $n_t = 0.34$ (magenta). The prediction for the simple slow-roll inflation model discussed in this section, $n_t = -r/8$, is shown in green. Figure taken from [107].

confidence limits on such a background are [111, 32, 158, 107]:

$$\begin{aligned} &\Omega_0 < 1.2 \times 10^{-9} \quad (\text{EPTA})\,, \\ &\Omega_0 < 8.5 \times 10^{-10} \quad (\text{NANOGrav})\,, \\ &\Omega_0 < 2.1 \times 10^{-10} \quad (\text{PPTA})\,. \end{aligned} \tag{10.20}$$

For a stochastic background from a population of black hole binaries on quasi-circular orbits driven by gravitational wave emission, $\beta = 2/3$. The most recent 95% confidence limits on such a background are [111, 32, 158]:

$$\begin{aligned} &\Omega_{2/3} < 5.4 \times 10^{-9} \quad (\text{EPTA})\,, \\ &\Omega_{2/3} < 1.3 \times 10^{-9} \quad (\text{NANOGrav})\,, \\ &\Omega_{2/3} < 6.0 \times 10^{-10} \quad (\text{PPTA})\,. \end{aligned} \tag{10.21}$$

10.2.3 Spacecraft Doppler tracking

Spacecraft Doppler tracking [30] operates on the same principles as pulsar timing, with a precision on-board clock and radio telemetry replacing the regular lighthouse-like radio emission of a pulsar. The ∼1–10 AU Earth-spacecraft separation places spacecraft Doppler tracking between pulsar timing and future LISA-like missions in terms of baseline and gravitational wave frequency coverage. In principle, a fleet of spacecraft each equipped with accurate clocks and high power radio transmitters could be used to perform the same type of cross-correlation analysis used in pulsar timing, but to-date the analyses have been limited to single spacecraft studies.

The most stringent bounds come from using the Cassini spacecraft, and place a bound on the strength of a stochastic gravitational-wave background at frequencies of order one over the transit time to the spacecraft [6]:

$$\Omega_{\rm gw}(f) < 0.027 \quad \text{for} \quad 10^{-6} < f < 10^{-3}\ \text{Hz}\,. \tag{10.22}$$

10.2.4 Interferometer bounds

Data from the initial LIGO and Virgo observation runs have been used to place constraints on the energy density of isotropic stochastic backgrounds across multiple frequency bands between 41.5 − 1726 Hz. The bounds are quoted in terms of $\Omega_{\rm gw}(f) = \Omega_\beta(f/f_{\rm ref})^\beta$ for $\beta = 0$ (flat in energy density) and $\beta = 3$ (flat in strain spectral density). The $\beta = 0$ bounds are quoted for the lower frequency bands, where the sensitivity is greatest for signals with this slope, while the $\beta = 3$ bounds are quoted for the higher frequency bands. The bounds assume a Hubble constant value of $H_0 = 68\ \text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}$.

Combining the initial LIGO and Virgo data, the most stringent 95% confidence limits for $\beta = 0$ are [2]:

$$\begin{aligned} &\Omega_{\rm gw}(f) < 5.6 \times 10^{-6} \quad \text{for} \quad 41.5 < f < 169.25\ \text{Hz}\,, \\ &\Omega_{\rm gw}(f) < 1.8 \times 10^{-4} \quad \text{for} \quad 170 < f < 600\ \text{Hz}\,. \end{aligned} \tag{10.23}$$

The bounds for $\beta = 3$ are [2, 4]:

$$\begin{aligned} \Omega_{\rm gw}(f) &< 7.7 \times 10^{-4} \left(\tfrac{f}{900\,\mathrm{Hz}}\right)^3 \quad \text{for} \quad 460 < f < 1000 \text{ Hz}\,, \\ \Omega_{\rm gw}(f) &< 1.0 \times 10^{-4} \left(\tfrac{f}{1300\,\mathrm{Hz}}\right)^3 \quad \text{for} \quad 1000 < f < 1726 \text{ Hz}\,. \end{aligned} \tag{10.24}$$

We note that the $\beta = 3$ bound for the $460 < f < 1000$ Hz frequency band comes from a correlation analysis using the colocated 2 km and 4 km Hanford detectors [4].

At the time of writing, the stochastic background analysis for the first advanced LIGO observation run is ongoing. The detection of multiple binary black hole mergers during this first observing run implies that stellar remnant black holes may produce a detectable stochastic signal from the superposition of many individually undetected sources [10]. This motivates searches for signals with the $\beta = 2/3$ spectral slope expected for such a population.

10.2.5 Bounds on anisotropic backgrounds

Constraints on anisotropic backgrounds have also been set using data from initial LIGO [5] and from the European Pulsar Timing Array [169]. The corresponding upper-limit maps are shown in Figures 84 and 85.

The upper-limit maps in the top two panels of Figure 84 were constructed using the spherical harmonic decomposition method (Section 7.3.6) for anisotropic backgrounds having spectral indices $\beta = 0$ and $\beta = 3$, and spherical harmonic components out to $l_{\rm max} = 7$ and 12, respectively. The upper limits for these maps are given in unit of $\text{strain}^2\,\mathrm{Hz}^{-1}\,\mathrm{sr}^{-1}$. The upper-limit map in the bottom panel of Figure 84 was constructed using the radiometer method (Section 7.3.6) for an anisotropic background having spectral index $\beta = 3$. (Recall that the radiometer method assumes a point-source signal model.) The units of this map are $\text{strain}^2\,\mathrm{Hz}^{-1}$. To express the upper limits for all the maps in units of $\text{strain}^2\,\mathrm{Hz}^{-1}$, one should multiply the top two maps by the solid angle $(\Delta\theta)^2$, where $\Delta\theta \simeq \pi/l_{\rm max}$; this is the angular resolution of the map constructed for spherical harmonic components out to $l_{\rm max}$. The reference frequency for all the maps was taken to be $f_{\rm ref} = 100$ Hz, which is contained in the bands 40–200 Hz and 40–500 Hz, which were analyzed for the $\beta = 0$ and $\beta = 3$ spectral indices, respectively.

The upper-limit map shown in Figure 85 is for the 2015 European Pulsar Timing Array data [169]. The map shows the 95% confidence-level upper limits on the (dimensionless) amplitude A_h of the characteristic strain (2.22):

$$h_c(f) = A_h \left(\frac{f}{\mathrm{yr}^{-1}}\right)^{-2/3} . \tag{10.25}$$

The spectral index $\alpha = -2/3$ is appropriate for a stochastic background formed from the superposition of gravitational-wave-driven, circular, inspiraling supermassive black-hole binaries, which is an expected source at the nano-Hz frequencies probed by pulsar timing arrays. The corresponding spectral index for the fractional energy density in gravitational waves, $\Omega_{\rm gw}(f)$, is $\beta = 2/3$ (Section 2.5).

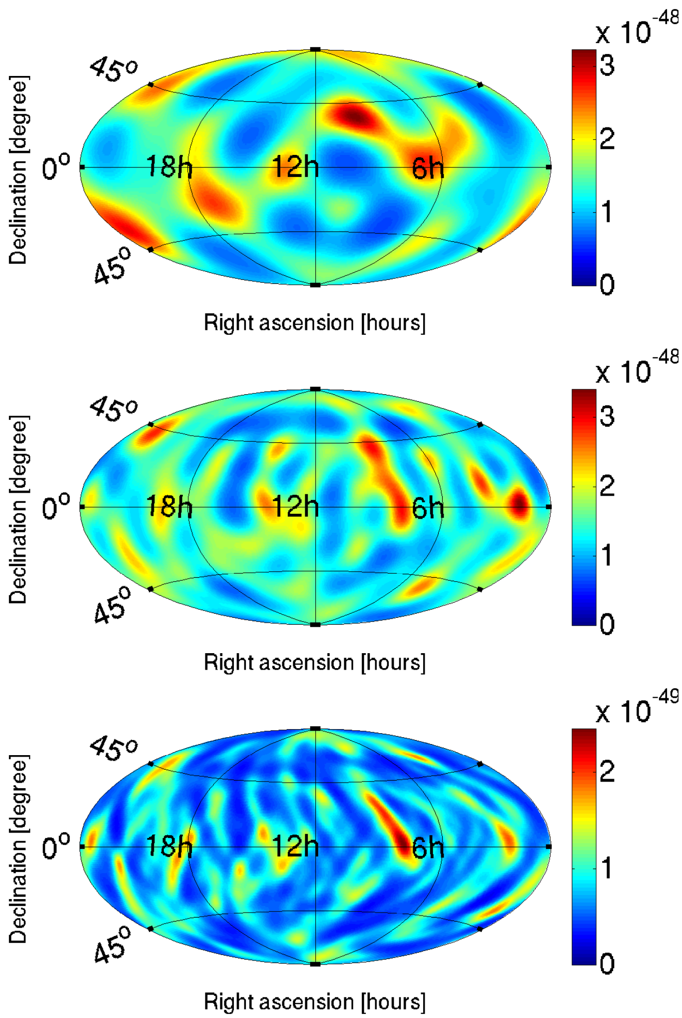


Figure 84: 90% confidence-level upper-limit maps on anisotropic backgrounds using initial LIGO data. Top panel: Upper-limit map on an anisotropic background having $\beta = 0$ and spherical harmonic components out to $l_{\rm max} = 7$, expressed in units of strain$^2\,{\rm Hz}^{-1}\,{\rm sr}^{-1}$. Middle panel: Same as the top panel but for an anisotropic background having $\beta = 3$ and spherical harmonic components out to $l_{\rm max} = 12$. Bottom panel: Upper limits on an anisotropic background having $\beta = 3$, constructed using the radiometer method, which assumes a point-source signal model. The units of this last map are strain$^2\,{\rm Hz}^{-1}$. Figure adapted from [5].

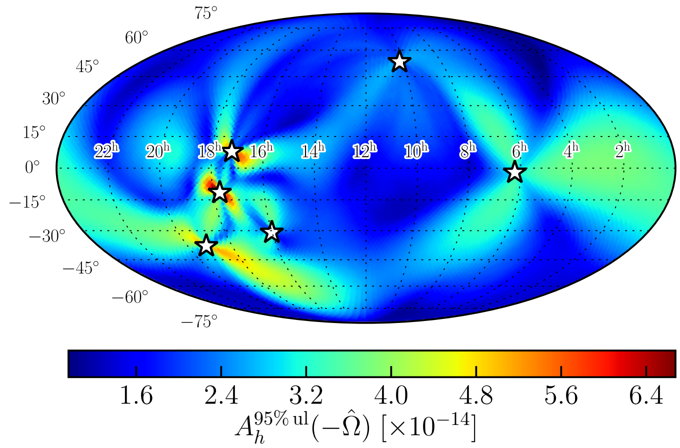


Figure 85: 95% confidence-level upper-limit map on the characteristic strain amplitude for an anisotropic background having spectral index $\alpha = -2/3$. The white stars show the location of the EPTA pulsars used for the analysis. Figure taken from [169].

11 Acknowledgements

JDR acknowledges support from National Science Foundation Awards PHY-1205585, CREST HRD-1242090, PHY-1505861. NJC acknowledges support from National Science Foundation Awards PHY-1306702 and PHY-1607343, and NASA award NNX16AB98G. JDR and NJC acknowledge support from the National Science Foundation NANOGrav Physics Frontier Center, NSF PFC-1430284. We also thank members of the LIGO-Virgo stochastic working group and members of NANOGrav for countless discussions related to all things stochastic. Special thanks go out to Bruce Allen, Matt Benacquista, Gwynne Crowder, Tyson Littenberg, Chris Messenger, Soumya Mohanty, Tanner Prestegard, Eric Thrane, and Michele Vallisneri, who either commented on parts of the text or provided figures for us to use. This research made use of Python and its standard libraries: numpy and matplotlib. We also made use of MEALPix (a Matlab implementation of HEALPix [76]), developed by the GWAstro Research Group and available from `http://gwastro.psu.edu`. Finally, we thank the editors of Living Reviews in Relativity (especially Bala Iyer and Frank Schulz) for their incredible patience while this article was being written. This document has been assigned LIGO Document Control Center number LIGO-P1600242.

A Freedom in the choice of polarization basis tensors

A.1 Linear polarization

In the main text, we chose the $A = +, \times$ polarization basis tensors to be

$$\begin{aligned} e^+_{ab}(\hat{n}) &= \hat{l}_a\hat{l}_b - \hat{m}_a\hat{m}_b\,, \\ e^\times_{ab}(\hat{n}) &= \hat{l}_a\hat{m}_b + \hat{m}_a\hat{l}_b\,, \end{aligned} \tag{A.1}$$

where $\hat{n}$ is the direction to the gravitational-wave source, and $\hat{l}$, $\hat{m}$ are unit vectors tangent to the sphere:

$$\begin{aligned} \hat{n} &= \sin\theta\cos\phi\,\hat{x} + \sin\theta\sin\phi\,\hat{y} + \cos\theta\,\hat{z} \equiv \hat{r}\,, \\ \hat{l} &= \cos\theta\cos\phi\,\hat{x} + \cos\theta\sin\phi\,\hat{y} - \sin\theta\,\hat{z} \equiv \hat{\theta}\,, \\ \hat{m} &= -\sin\phi\,\hat{x} + \cos\phi\,\hat{y} \equiv \hat{\phi}\,. \end{aligned} \tag{A.2}$$

This particular choice for the vectors $\hat{l}$, $\hat{m}$, perpendicular to $\hat{n}$ is somewhat arbitrary, as one can rotate these vectors by an angle ψ in the plane orthogonal to $\hat{n}$, preserving the triple as a right-handed orthonormal triad. (For a gravitational-wave source with a symmetry axis, such as a binary system or rotating neutron star, the angle ψ can be interpreted as the *polarization angle* of the source.) See Figure 86. Under such a rotation, $\hat{l}$ and $\hat{m}$ transform to new unit vectors

$$\begin{aligned} \hat{p} &\equiv \cos\psi\,\hat{l} + \sin\psi\,\hat{m}\,, \\ \hat{q} &\equiv -\sin\psi\,\hat{l} + \cos\psi\,\hat{m}\,, \end{aligned} \tag{A.3}$$

leading to new polarization tensors

$$\begin{aligned} \epsilon^+_{ab}(\hat{n},\psi) &\equiv \hat{p}_a\hat{p}_b - \hat{q}_a\hat{q}_b\,, \\ \epsilon^\times_{ab}(\hat{n},\psi) &\equiv \hat{p}_a\hat{q}_b + \hat{q}_a\hat{p}_b\,. \end{aligned} \tag{A.4}$$

The new polarization tensors are related to the original ones via

$$\begin{aligned} \epsilon^+_{ab}(\hat{n},\psi) &= \cos 2\psi\, e^+_{ab}(\hat{n}) + \sin 2\psi\, e^\times_{ab}(\hat{n})\,, \\ \epsilon^\times_{ab}(\hat{n},\psi) &= -\sin 2\psi\, e^+_{ab}(\hat{n}) + \cos 2\psi\, e^\times_{ab}(\hat{n})\,. \end{aligned} \tag{A.5}$$

A.2 Circular polarization

The form of the above transformation suggests a more convenient basis of polarization tensors. Namely, if we define the complex combinations

$$\begin{aligned} e^R_{ab} &\equiv \frac{1}{\sqrt{2}}\left(e^+_{ab} + i\,e^\times_{ab}\right)\,, \\ e^L_{ab} &\equiv \frac{1}{\sqrt{2}}\left(e^+_{ab} - i\,e^\times_{ab}\right)\,, \end{aligned} \tag{A.6}$$

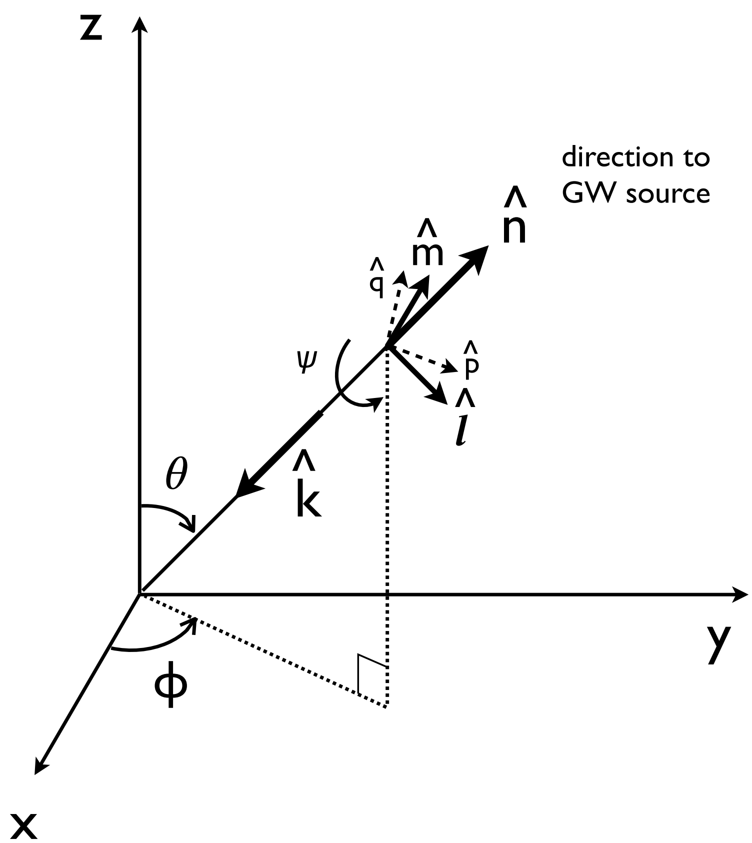


Figure 86: Different choices for the unit vectors perpendicular to $\hat{n}$. By rotating the unit vectors $\hat{l}$, $\hat{m}$ by the angle ψ in the plane orthogonal to $\hat{n}$, one obtains new unit vectors, $\hat{p}$, $\hat{q}$, in terms of which new polarization basis tensors, $\epsilon^{+}_{ab}(\hat{n},\psi)$, $\epsilon^{\times}_{ab}(\hat{n},\psi)$, are defined.

or, equivalently,

$$
\begin{aligned}
e^R_{ab} &\equiv \frac{1}{\sqrt{2}}(\hat{l}_a + i\hat{m}_a)(\hat{l}_b + i\hat{m}_b)\,,\\
e^L_{ab} &\equiv \frac{1}{\sqrt{2}}(\hat{l}_a - i\hat{m}_a)(\hat{l}_b - i\hat{m}_b)\,,
\end{aligned}
\tag{A.7}
$$

then under the above rotation by ψ,

$$
\begin{aligned}
\epsilon^R_{ab}(\hat{n},\psi) &= e^{-i2\psi}\, e^R_{ab}(\hat{n})\,,\\
\epsilon^L_{ab}(\hat{n},\psi) &= e^{i2\psi}\, e^L_{ab}(\hat{n})\,.
\end{aligned}
\tag{A.8}
$$

The tensors e^R_{ab}, e^L_{ab} correspond to *right* and *left* circularly polarized waves when looking down on the $\{\hat{l},\hat{m}\}$ plane in the $-\hat{n}$-direction. (The deformation ellipse for e^R_{ab} would rotate to the right, i.e., clockwise, when viewed in this direction.) The fact that the right and left circularly polarized waves transform by a simple phase factor involving 2ψ is a manifestation of the spin-two nature of the graviton [196]. Indeed, one can show that the scalar field $e^R_{ab}(\hat{n})h^{ab}(f,\hat{n})$ can be written as a linear combination of spin-weight $+2$ spherical harmonics ${}_2Y_{lm}(\hat{n})$, while $e^L_{ab}(\hat{n})h^{ab}(f,\hat{n})$ can be written as a linear combination of spin-weight -2 spherical harmonics ${}_{-2}Y_{lm}(\hat{n})$. (See Appendices D, E, F for more details regarding spin-weighted and vector and tensor spherical harmonics.)

The Fourier components $h_{ab}(f,\hat{n})$ of the metric perturbations $h_{ab}(t,\vec{x})$ can be expanded in terms of either the linear polarization basis tensors:

$$
h_{ab}(f,\hat{n}) = h_+(f,\hat{n})e^+_{ab}(\hat{n}) + h_\times(f,\hat{n})e^\times_{ab}(\hat{n})\,,
\tag{A.9}
$$

or the circular polarization basis tensors:

$$
h_{ab}(f,\hat{n}) = h_R(f,\hat{n})e^R_{ab}(\hat{n}) + h_L(f,\hat{n})e^L_{ab}(\hat{n})\,.
\tag{A.10}
$$

The expansion coefficients h_R, h_L are related to h_+, $h_\times$ via:

$$
\begin{aligned}
h_R &= \frac{1}{\sqrt{2}}\left(h_+ - ih_\times\right)\,,\\
h_L &= \frac{1}{\sqrt{2}}\left(h_+ + ih_\times\right)\,.
\end{aligned}
\tag{A.11}
$$

Note the sign change on the right-hand side of (A.11) compared to (A.6).

A.3 Polarization matrix and Stokes' parameters

For a single monochromatic plane wave, the expansion coefficients h_+, $h_\times$ or h_R, h_L are (complex-valued) constants. The polarization content of the plane wave is encoded in terms of the 2×2 (Hermitian) polarization matrix

$$
J_{BB'} \equiv h_B h^*_{B'}\,,
\tag{A.12}
$$

where B labels either the linear polarization components $A \equiv \{+, \times\}$ or circular polarization components $C \equiv \{R, L\}$. For linear polarization, the matrix elements have the form

$$J_{AA'} = \frac{1}{2}\left[\begin{array}{cc} I+Q & U-iV \\ U+iV & I-Q \end{array}\right], \tag{A.13}$$

where I, Q, U, V are the *Stokes' parameters* [93]:

$$\begin{aligned} I &= |h_+|^2 + |h_\times|^2 \,, \\ Q &= |h_+|^2 - |h_\times|^2 \,, \\ U &= h_+ h_\times^* + h_\times h_+^* \,, \\ V &= i(h_+ h_\times^* - h_\times h_+^*) \,. \end{aligned} \tag{A.14}$$

For circular polarization, we have

$$J_{CC'} = \frac{1}{2}\left[\begin{array}{cc} I+V & Q-iU \\ Q+iU & I-V \end{array}\right], \tag{A.15}$$

where

$$\begin{aligned} I &= |h_R|^2 + |h_L|^2 \,, \\ Q &= h_R h_L^* + h_L h_R^* \,, \\ U &= i(h_R h_L^* - h_L h_R^*) \,, \\ V &= |h_R|^2 - |h_L|^2 \,. \end{aligned} \tag{A.16}$$

Note that I is the total intensity of the wave, Q is a measure of linear polarization, $|h_+|^2 - |h_\times|^2$, and V is a measure of circular polarization, $|h_R|^2 - |h_L|^2$. Since a stochastic gravitational-wave background is a linear *superposition* of plane waves having different frequencies and coming from different directions on the sky, the matrix elements of J will be replaced by quadractic *expectation values*, e.g., $\langle h_+(f, \hat{n}) h_\times^*(f', \hat{n}')\rangle$, which will also depend on whether the background is stationary or anisotropic, etc.

Given the transformation properties (A.8) of e^R_{ab}, e^L_{ab}, and the definition (A.11) of h_R, h_L, it follows that h_R, h_L transform to

$$\begin{aligned} \bar{h}_R &= e^{i2\psi} h_R \,, \\ \bar{h}_L &= e^{-i2\psi} h_L \,, \end{aligned} \tag{A.17}$$

under a rotation of the basis vectors $\{\hat{l}, \hat{m}\}$ by ψ. From these equations and expressions (A.16) for the Stokes parameters, we can further show that I, Q, U, V transform to

$$\begin{aligned} \bar{I} &= I \,, \\ \bar{V} &= V \,, \\ \bar{Q} + i\bar{U} &= e^{-i4\psi}(Q + iU) \,, \\ \bar{Q} - i\bar{U} &= e^{i4\psi}(Q - iU) \,, \end{aligned} \tag{A.18}$$

under a rotation by ψ. Thus, I and V are ordinary scalar (spin 0) functions on the sphere, while $Q \pm iU$ are spin 4 fields, and can be written as linear combinations of spin-weight ± 4

spherical harmonics ${}_{\pm 4}Y_{lm}(\hat{n})$. This has relevance for searches for circularly or linearly polarized stochastic backgrounds, as circular polarization, V, is present in the isotropic component of the background, while linear polarization, Q, is not [152].

B Some standard results for Gaussian random variables

The statistical properties of a random variable X are completely determined by its probability distribution $p_X(x)$. The *moments* of the distribution $\langle X\rangle$, $\langle X^2\rangle$, $\langle X^3\rangle, \cdots$, are defined by

$$\langle X^n\rangle \equiv \int_{-\infty}^{\infty} dx\, x^n p_X(x)\,. \tag{B.1}$$

The first moment $\langle X\rangle$ is the expected (or mean) value of X, and is often denoted by μ; the second moment is related to the variance σ^2 via the formula $\langle X^2\rangle = \sigma^2 + \langle X\rangle^2$. The *characteristic function* of the probability distribution is defined by the Fourier transform:

$$\varphi_X(t) \equiv \int_{-\infty}^{\infty} dx\, e^{itx}\, p_X(x)\,. \tag{B.2}$$

Note that by expanding the exponential

$$\varphi_X(t) = 1 + it\langle X\rangle + \frac{i^2t^2}{2!}\langle X^2\rangle + \cdots\,. \tag{B.3}$$

This means that the moments $\langle X^n\rangle$ can be obtained by simply differentiating $\varphi_X(t)$:

$$\langle X^n\rangle = i^{-n}\left[\frac{d^n}{dt^n}\varphi_X(t)\right]\Bigg|_{t=0}\,. \tag{B.4}$$

If the moments are all finite and the expansion (B.3) is absolutely convergent near the origin, then the probability distribution $p_X(x)$ is simply the inverse Fourier transform of $\varphi_X(t)$:

$$p_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dt\, e^{-itx}\,\varphi_X(t)\,. \tag{B.5}$$

A similar result can be obtained for a *one-sided* probability distribution $p_X(x)$ (e.g., defined only for $x \geq 0$) by working with Laplace transformations instead.

If X is a *Gaussian* random variable, then

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}\,. \tag{B.6}$$

The parameters μ and σ^2 are just the mean and variance of X:

$$\mu = \langle X\rangle\,, \qquad \sigma^2 = \langle X^2\rangle - \langle X\rangle^2\,. \tag{B.7}$$

A nice property of Gaussian distributions is that all third and higher-order moments can be expressed as a sum of products of the first two moments. For example, for a single Gaussian random variable X,

$$\begin{aligned}
\langle X^3\rangle &= 3\langle X\rangle\langle X^2\rangle - 2\langle X\rangle^3\\
\langle X^4\rangle &= 4\langle X\rangle\langle X^3\rangle + 3\langle X^2\rangle^2 - 12\langle X\rangle^2\langle X^2\rangle + 6\langle X\rangle^4\\
&\cdots
\end{aligned} \tag{B.8}$$

More generally, for $n \geq 3$ these relations can be obtained by solving the equations

$$\left[\frac{d^n}{dt^n} \ln \varphi_X(t)\right]\Bigg|_{t=0} = 0 \tag{B.9}$$

for $\langle X^n \rangle$, where $\varphi_X(t)$ is given by the right-hand side of (B.3). The fact that the derivatives are actually equal to zero follows from the specific form for the characteristic function for a Gaussian distribution:

$$\varphi_X(t) = \exp\left[i\mu t - \frac{\sigma^2 t^2}{2}\right] . \tag{B.10}$$

Since $\ln \varphi_X(t)$ is quadratic in t, all third and higher order derivatives vanish.

A *multivariate Gaussian* distribution is a generalization of (B.6) to a set of random variable $\mathbf{X} \equiv \{X_1, X_2, \cdots, X_N\}$. The joint probability density function is given by

$$p_{\mathbf{X}}(x_1, x_2, \cdots, x_N) = \frac{1}{\sqrt{\det(2\pi C)}}\, e^{-\frac{1}{2}\sum_{i,j}(x_i-\mu_i)C^{-1}_{ij}(x_j-\mu_j)} , \tag{B.11}$$

where $\mu_i = \langle X_i \rangle$ are the mean values, and

$$C_{ij} = \langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle \tag{B.12}$$

are the elements of the *covariance matrix* C. For a zero-mean multivariate Gaussian distribution, all of the odd-ordered moments are identically zero. In addition,

$$\langle X_1 X_2 X_3 X_4 \rangle = \langle X_1 X_2 \rangle \langle X_3 X_4 \rangle + \langle X_1 X_3 \rangle \langle X_2 X_4 \rangle + \langle X_1 X_4 \rangle \langle X_2 X_3 \rangle . \tag{B.13}$$

We will use several of the above results repeatedly throughout the main text, as most of the probability distributions that we work with are multivariate-Gaussian.

C Definitions and tests for stationarity and Gaussianity

Here we provide definitions of what it means for data to be stationary and Gaussian, and highlight some tests for these properties. Ascertaining whether or not data are stationary and Gaussian can be challenging as the tests rely on comparison with alternative models, and some models are better at picking up certain forms of non-stationarity and non-Gaussianity than others.

C.1 Definition of stationarity

A stationary stochastic process has statistical properties that do not depend on time: the joint statistical properties of the sample $\{x_{t_1}, \ldots, x_{t_k}\}$ are identical to the joint statistical properties of the sample $\{x_{t_1+\tau}, \ldots, x_{t_k+\tau}\}$ for all τ and k. In particular, the joint distribution of (x_t, x_s) depends only on the lag $t-s$, and not on t or s, and all higher moments are strictly independent of time. A less restrictive, and more practical notion, is that of *weak or second-order stationarity*, which asserts that the mean and variance are constant, and that the auto-covariance $\mathrm{cov}(x_t, x_{t+\tau})$ depends only on the lag τ.

C.2 Definition of Gaussian

A continuous random variable X is said to be a Gaussian, or normal, random variable $X \sim N(\mu, \sigma^2)$ if its probability density function is given by

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} . \tag{C.1}$$

The multivariate generalization to a collection of continuous random variables $\mathbf{X} \equiv \{X_1, X_2, \cdots, X_N\}$ is given in terms of a Gaussian probability density function with covariance matrix C:

$$p_{\mathbf{X}}(x_1, x_2, \cdots, x_N) = \frac{1}{\sqrt{\det(2\pi C)}} e^{-\frac{1}{2}\sum_{i,j}(x_i-\mu_i)C_{ij}^{-1}(x_j-\mu_j)} . \tag{C.2}$$

See Appendix B for additional statistical properties of Gaussian random variables.

C.3 Tests for stationarity

There exists a vast literature on tests of non-stationarity of time-series data. The simplest tests for non-stationarity are qualitative in nature and involve looking at plots of the mean, variance, and auto-correlation as a function of time (for example, by using a sliding window of some duration to select the samples used to compute these quantities). The difficulty with this approach is deciding on what constitutes acceptable levels of variation. The concept of time-varying correlations and time-varying spectral densities are well defined and useful concepts for *locally-stationary* processes [55], but less so for other forms of non-stationarity.

It is unclear whether many of the more powerful quantitative tests for non-stationarity are useful for gravitational-wave data analysis. For example, commonly used tests, such as the augmented Dickey-Fuller test and the Phillips-Perron test, which test to see if the data follow a "unit root" auto-regressive process, do not appear to be particularly applicable since the noise encountered in gravitational-wave experiments usually exhibits high auto-correlation, and thus has roots that are naturally close to unity, which poses a challenge for these tests [126].

The most useful tests, at least for evenly sampled gravitational-wave data, are those based on evolutionary spectral estimates, or correlations in the Fourier coefficients. The Priestley-Subba Rao test [114], and modern variants based on wavelets [127], use window functions to compute spectral estimates as a function of time. A statistical test is then used to assess if the spectral estimates are consistent with stationarity. The second type of test is based on the fact that second-order stationary time series produce uncorrelated Fourier series (which is why most gravitational-wave analyses are performed in the Fourier domain). Statistical tests can be used to decide whether the the level of correlation between Fourier coefficients indicates that the data are non-stationary [59].

C.4 Tests for Gaussianity

There are a large number of tests for Gaussianity described in the literature that are in regular use. These tests are based on different properties of the Gaussian distribution, and the power of the tests differ depending on the nature of the non-Gaussianity.

Three of the most widely used frequentist tests are the Shapiro–Wilk test, the Anderson–Darling test and the Lilliefors test (a modified Kolmogorov–Smirnov test). The Shapiro–Wilk test is a regression test that out-performs other tests on small data sets, but is challenging to apply to the large data sets encountered in gravitational-wave data analysis. Both the Anderson–Darling and the Lilliefors test are based on the distance between the hypothesized cumulative distribution function (in this case, that of a Gaussian distribution) and the cumulative distribution function of the data. The Anderson–Darling test performs almost as well, and sometimes better, than the Shapiro–Wilk test [149], and can be used on large data sets.

Bayesian tests for Gaussianity can be performed by computing the Bayes Factors between competing models for the data, in this case the Gaussian distribution and some more general alternative such as the Student's t-distribution [162, 104]. This approach has been applied in gravitational-wave data analysis [112, 44].

D Ordinary (scalar) and spin-weighted spherical harmonics

This appendix, adapted from [71], summarizes some useful relations involving spin-weighted and ordinary spherical harmonics, ${}_sY_{lm}(\hat{n})$ and $Y_{lm}(\hat{n})$. For more details, see e.g., [75] and [56]. Note that for our analyses, we can restrict attention to spin-weighted spherical harmonics having *integral* spin weight s, even though spin-weighted spherical harmonics with half-integral spin weight do exist.

Ordinary spherical harmonics:

$$Y_{lm}(\hat{n}) \equiv Y_{lm}(\theta,\phi) = N_l^m P_l^m(\cos\theta) e^{im\phi}\,, \quad \text{where } N_l^m = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}\,. \tag{D.1}$$

Relation of spin-weighted spherical harmonics to ordinary spherical harmonics:

$$\begin{aligned} {}_sY_{lm}(\theta,\phi) &= \sqrt{\frac{(l-s)!}{(l+s)!}}\,\eth^s Y_{lm}(\theta,\phi) \quad \text{for} \quad 0 \le s \le l\,, \\ {}_sY_{lm}(\theta,\phi) &= \sqrt{\frac{(l+s)!}{(l-s)!}}\,(-1)^s \bar{\eth}^{-s} Y_{lm}(\theta,\phi) \quad \text{for} \quad -l \le s \le 0\,, \end{aligned} \tag{D.2}$$

where

$$\begin{aligned} \eth\eta &= -(\sin\theta)^s \left[\frac{\partial}{\partial\theta} + i\csc\theta\frac{\partial}{\partial\phi}\right](\sin\theta)^{-s}\eta\,, \\ \bar{\eth}\eta &= -(\sin\theta)^{-s}\left[\frac{\partial}{\partial\theta} - i\csc\theta\frac{\partial}{\partial\phi}\right](\sin\theta)^s\eta\,, \end{aligned} \tag{D.3}$$

and $\eta = \eta(\theta,\phi)$ is a spin-s scalar field.

Complex conjugate:

$${}_sY^*_{lm}(\theta,\phi) = (-1)^{m+s}\,{}_{-s}Y_{l,-m}(\theta,\phi)\,. \tag{D.4}$$

Relation to Wigner rotation matrices:

$$D^l{}_{m'm}(\phi,\theta,\psi) = (-1)^{m'}\sqrt{\frac{4\pi}{2l+1}}\,{}_mY_{l,-m'}(\theta,\phi)e^{-im\psi}\,, \tag{D.5}$$

or

$$\left[D^l{}_{m'm}(\phi,\theta,\psi)\right]^* = (-1)^m\sqrt{\frac{4\pi}{2l+1}}\,{}_{-m}Y_{l,m'}(\theta,\phi)e^{im\psi}\,. \tag{D.6}$$

Parity transformation:

$${}_sY_{lm}(\pi-\theta,\phi+\pi) = (-1)^l\,{}_{-s}Y_{lm}(\theta,\phi)\,. \tag{D.7}$$

Orthonormality (for fixed s):

$$\int d^2\Omega_{\hat{n}}\,{}_sY_{lm}(\hat{n})\,{}_sY^*_{l'm'}(\hat{n}) \equiv \int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta\, d\theta\,{}_sY_{lm}(\theta,\phi)\,{}_sY^*_{l'm'}(\theta,\phi) = \delta_{ll'}\delta_{mm'}\,. \tag{D.8}$$

Addition theorem for spin-weighted spherical harmonics:

$$\sum_{m=-l}^{l} {}_sY_{lm}(\theta_1,\phi_1)\, {}_{s'}Y^*_{lm}(\theta_2,\phi_2) = (-1)^{-s'}\sqrt{\frac{2l+1}{4\pi}}\, {}_{-s'}Y_{ls}(\theta_3,\phi_3)e^{is'\chi_3}\,, \tag{D.9}$$

where

$$\cos\theta_3 = \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2\cos(\phi_2-\phi_1)\,, \tag{D.10}$$

and

$$\begin{aligned} e^{-i(\phi_3+\chi_3)/2} &= \frac{\cos\frac{1}{2}(\phi_2-\phi_1)\cos\frac{1}{2}(\theta_2-\theta_1) - i\sin\frac{1}{2}(\phi_2-\phi_1)\cos\frac{1}{2}(\theta_1+\theta_2)}{\sqrt{\cos^2\frac{1}{2}(\phi_2-\phi_1)\cos^2\frac{1}{2}(\theta_2-\theta_1)+\sin^2\frac{1}{2}(\phi_2-\phi_1)\cos^2\frac{1}{2}(\theta_1+\theta_2)}}\,, \\ e^{i(\phi_3-\chi_3)/2} &= \frac{\cos\frac{1}{2}(\phi_2-\phi_1)\sin\frac{1}{2}(\theta_2-\theta_1) + i\sin\frac{1}{2}(\phi_2-\phi_1)\sin\frac{1}{2}(\theta_1+\theta_2)}{\sqrt{\cos^2\frac{1}{2}(\phi_2-\phi_1)\sin^2\frac{1}{2}(\theta_2-\theta_1)+\sin^2\frac{1}{2}(\phi_2-\phi_1)\sin^2\frac{1}{2}(\theta_1+\theta_2)}}\,. \end{aligned} \tag{D.11}$$

Addition theorem for ordinary spherical harmonics:

$$\sum_{m=-l}^{l} Y_{lm}(\hat{n}_1)Y^*_{lm}(\hat{n}_2) = \frac{2l+1}{4\pi}\,P_l(\hat{n}_1\cdot\hat{n}_2)\,. \tag{D.12}$$

Integral of a product of spin-weighted spherical harmonics:

$$\begin{aligned} &\int d^2\Omega_{\hat{n}}\ {}_{s_1}Y_{l_1m_1}(\hat{n})\, {}_{s_2}Y_{l_2m_3}(\hat{n})\, {}_{s_3}Y_{l_3m_3}(\hat{n}) \\ &\quad= \sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}}\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3\end{pmatrix}\begin{pmatrix} l_1 & l_2 & l_3 \\ -s_1 & -s_2 & -s_3\end{pmatrix}\,, \end{aligned} \tag{D.13}$$

where $\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3\end{pmatrix}$ is a Wigner 3-j symbol [197, 118].

E Gradient and curl rank-1 (vector) spherical harmonics

The gradient and curl rank-1 (vector) spherical harmonics are defined for $l \geq 1$ by:

$$
\begin{aligned}
Y^G_{(lm)a} &\equiv \frac{1}{2}{}^{(1)}N_l \partial_a Y_{lm} = \frac{1}{2}{}^{(1)}N_l \left(\frac{\partial Y_{lm}}{\partial \theta}\,\hat{\theta}_a + \frac{1}{\sin\theta}\frac{\partial Y_{lm}}{\partial \phi}\,\hat{\phi}_a \right), \\
Y^C_{(lm)a} &\equiv \frac{1}{2}{}^{(1)}N_l (\partial_b Y_{lm}) \epsilon^b{}_a = \frac{1}{2}{}^{(1)}N_l \left(-\frac{1}{\sin\theta}\frac{\partial Y_{lm}}{\partial \phi}\,\hat{\theta}_a + \frac{\partial Y_{lm}}{\partial \theta}\,\hat{\phi}_a \right),
\end{aligned}
\tag{E.1}
$$

where $\hat{\theta}$ and $\hat{\phi}$ are the standard unit vectors tangent to the 2-sphere

$$
\begin{aligned}
\hat{\theta} &= \cos\theta\cos\phi\,\hat{x} + \cos\theta\sin\phi\,\hat{y} - \sin\theta\,\hat{z}\,, \\
\hat{\phi} &= -\sin\phi\,\hat{x} + \cos\phi\,\hat{y}\,,
\end{aligned}
\tag{E.2}
$$

${}^{(1)}N_l$ is a normalisation constant

$$
{}^{(1)}N_l = \sqrt{\frac{2(l-1)!}{(l+1)!}}\,,
\tag{E.3}
$$

and ϵ_{ab} is the Levi-Civita anti-symmetric tensor

$$
\epsilon_{ab} = \sqrt{g}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad g \equiv \det(g_{ab})\,.
\tag{E.4}
$$

Following standard practice, we use the metric tensor g_{ab} on the 2-sphere and its inverse g^{ab} to "lower" and "raise" tensor indices—e.g., $\epsilon^c{}_b \equiv g^{ca}\epsilon_{ab}$. In standard spherical coordinates (θ, ϕ),

$$
g_{ab} = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{pmatrix}, \qquad \sqrt{g} = \sin\theta\,.
\tag{E.5}
$$

The gradient and curl spherical harmonics are related to the spin-weight ± 1 spherical harmonics

$$
{}_{\pm 1}Y_{lm}(\theta,\phi) = \sqrt{\frac{(l-1)!}{(l+1)!}}\frac{N^m_l}{\sqrt{1-x^2}}\left(\pm(1-x^2)\frac{dP^m_l}{dx} + mP^m_l(x) \right) \mathrm{e}^{im\phi}\,,
\tag{E.6}
$$

where $x = \cos\theta$, via

$$
Y^G_{(lm)a} \pm iY^C_{(lm)a} = \pm\frac{1}{\sqrt{2}}(\hat{\theta}_a \pm i\hat{\phi}_a)\,{}_{\mp 1}Y_{lm}\,,
\tag{E.7}
$$

or, equivalently,

$$
\begin{aligned}
Y^G_{(lm)a} &= \frac{1}{2\sqrt{2}}\left[({}_{-1}Y_{lm} - {}_{1}Y_{lm})\,\hat{\theta}_a + i\,({}_{-1}Y_{lm} + {}_{1}Y_{lm})\,\hat{\phi}_a \right], \\
Y^C_{(lm)a} &= \frac{1}{2\sqrt{2}}\left[({}_{-1}Y_{lm} - {}_{1}Y_{lm})\,\hat{\phi}_a - i\,({}_{-1}Y_{lm} + {}_{1}Y_{lm})\,\hat{\theta}_a \right].
\end{aligned}
\tag{E.8}
$$

For decompositions of vector-longitudinal backgrounds, as discussed in the main text, it will be convenient to construct rank-2 tensor fields

$$\begin{aligned} Y^{V_G}_{(lm)ab} &= Y^{G}_{(lm)a}\hat{n}_b + Y^{G}_{(lm)b}\hat{n}_a \,, \\ Y^{V_C}_{(lm)ab} &= Y^{C}_{(lm)a}\hat{n}_b + Y^{C}_{(lm)b}\hat{n}_a \,, \end{aligned} \tag{E.9}$$

where $\hat{n}$ is the unit radial vector orthogonal to the surface of the 2-sphere:

$$\hat{n} = \sin\theta\cos\phi\,\hat{x} + \sin\theta\sin\phi\,\hat{y} + \cos\theta\,\hat{z}\,. \tag{E.10}$$

These fields satisfy the following orthonormality relations

$$\begin{aligned} \int d^2\Omega_{\hat{n}}\; Y^{V_G}_{(lm)ab}(\hat{n}) Y^{V_G}_{(l'm')}{}^{ab\,*}(\hat{n}) &= \delta_{ll'}\delta_{mm'} \,, \\ \int d^2\Omega_{\hat{n}}\; Y^{V_C}_{(lm)ab}(\hat{n}) Y^{V_C}_{(l'm')}{}^{ab\,*}(\hat{n}) &= \delta_{ll'}\delta_{mm'} \,, \\ \int d^2\Omega_{\hat{n}}\; Y^{V_G}_{(lm)ab}(\hat{n}) Y^{V_C}_{(l'm')}{}^{ab\,*}(\hat{n}) &= 0\,. \end{aligned} \tag{E.11}$$

F Gradient and curl rank-2 (tensor) spherical harmonics

The gradient and curl rank-2 (tensor) spherical harmonics are defined for $l \geq 2$ by:

$$\begin{aligned} Y^G_{(lm)ab} &= {}^{(2)}N_l \left(Y_{(lm);ab} - \frac{1}{2} g_{ab} Y_{(lm);c}{}^{c} \right) , \\ Y^C_{(lm)ab} &= \frac{{}^{(2)}N_l}{2} \left(Y_{(lm);ac} \epsilon^c{}_b + Y_{(lm);bc} \epsilon^c{}_a \right) , \end{aligned} \tag{F.1}$$

where a semicolon denotes covariant derivative on the 2-sphere, and ${}^{(2)}N_l$ is a normalisation constant

$${}^{(2)}N_l = \sqrt{\frac{2(l-2)!}{(l+2)!}} . \tag{F.2}$$

Using the standard polarization tensors on the 2-sphere:

$$\begin{aligned} e^+_{ab}(\hat{n}) &= \hat{\theta}_a \hat{\theta}_b - \hat{\phi}_a \hat{\phi}_b , \\ e^\times_{ab}(\hat{n}) &= \hat{\theta}_a \hat{\phi}_b + \hat{\phi}_a \hat{\theta}_b , \end{aligned} \tag{F.3}$$

where $\hat{\theta}$, $\hat{\phi}$ are given by (E.2) and $\hat{n}$ by (E.10), we have [90]:

$$\begin{aligned} Y^G_{(lm)ab}(\hat{n}) &= \frac{{}^{(2)}N_l}{2} \left[W_{(lm)}(\hat{n}) e^+_{ab}(\hat{n}) + X_{(lm)}(\hat{n}) e^\times_{ab}(\hat{n}) \right] , \\ Y^C_{(lm)ab}(\hat{n}) &= \frac{{}^{(2)}N_l}{2} \left[W_{(lm)}(\hat{n}) e^\times_{ab}(\hat{n}) - X_{(lm)}(\hat{n}) e^+_{ab}(\hat{n}) \right] , \end{aligned} \tag{F.4}$$

where

$$\begin{aligned} W_{(lm)}(\hat{n}) &= \left(\frac{\partial^2}{\partial\theta^2} - \cot\theta \frac{\partial}{\partial\theta} + \frac{m^2}{\sin^2\theta} \right) Y_{lm}(\hat{n}) = \left(2\frac{\partial^2}{\partial\theta^2} + l(l+1) \right) Y_{lm}(\hat{n}) , \\ X_{(lm)}(\hat{n}) &= \frac{2im}{\sin\theta} \left(\frac{\partial}{\partial\theta} - \cot\theta \right) Y_{lm}(\hat{n}) . \end{aligned} \tag{F.5}$$

These functions enter the expression for the spin-weight ± 2 spherical harmonics [129, 75]:

$${}_{\pm 2}Y_{lm}(\hat{n}) = \frac{{}^{(2)}N_l}{\sqrt{2}} \left[W_{(lm)}(\hat{n}) \pm i X_{(lm)}(\hat{n}) \right] , \tag{F.6}$$

which are related to the gradient and curl spherical harmonics via

$$Y^G_{(lm)ab}(\hat{n}) \pm i Y^C_{(lm)ab}(\hat{n}) = \frac{1}{\sqrt{2}} \left(e^+_{ab}(\hat{n}) \pm i e^\times_{ab}(\hat{n}) \right) {}_{\mp 2}Y_{lm}(\hat{n}) . \tag{F.7}$$

Note that the gradient and curl spherical harmonics satisfy the orthonormality relations

$$\begin{aligned} &\int_{S^2} d^2\Omega_{\hat{n}} \, Y^G_{(lm)ab}(\hat{n}) Y^{G\,ab\,*}_{(l'm')}(\hat{n}) = \delta_{ll'} \delta_{mm'} , \\ &\int_{S^2} d^2\Omega_{\hat{n}} \, Y^C_{(lm)ab}(\hat{n}) Y^{C\,ab\,*}_{(l'm')}(\hat{n}) = \delta_{ll'} \delta_{mm'} , \\ &\int_{S^2} d^2\Omega_{\hat{n}} \, Y^G_{(lm)ab}(\hat{n}) Y^{C\,ab\,*}_{(l'm')}(\hat{n}) = 0 . \end{aligned} \tag{F.8}$$

G Translation between $\hat{n}$ and $\hat{k}$ conventions

Numerous papers on detecting stochastic gravitational-wave backgrounds have adopted the convention where the polarization tensors and detector response functions are functions of the *direction of propagation* of the gravitational wave, $\hat{k}$, where $\hat{k}$ points radially *outward*. In this article, we have adopted instead the convention where plane wave expansions, polarization tensors, and response functions are written in terms of the *direction to the source* of the gravitational wave, $\hat{n}$, where again $\hat{n}$ points radially outward. In both approaches, the unit vectors $\hat{l}$ and $\hat{m}$, which are perpendicular to $\hat{k}$ (or $\hat{n}$) and are used to define the polarization tensors, are typically chosen to be the standard spherical polar coordinate unit vectors $\hat{\theta}$ and $\hat{\phi}$. Thus, the polarization tensors $e_{ab}^{+,\times}(\hat{n})$ and $e_{ab}^{+,\times}(\hat{k})$ are the same for both conventions. What is different is the expression for an individual plane wave—either $e^{i2\pi f(t-\hat{k}\cdot\vec{x}/c)}$ or $e^{i2\pi f(t+\hat{n}\cdot\vec{x}/c)}$—as the direction of propagation of the wave is opposite the direction to the source.

In this appendix, we summarize how the expressions for the response functions $R^{ab}(f,\hat{n})$, $R^{A}(f,\hat{n})$, and $R^{P}_{(lm)}(f)$, given in previous sections are related to similar quantities calculated in other papers that use the $\hat{k}$-convention. For completeness, we will write down expressions for the vector and scalar polarization modes (Section 8.3) in addition to the standard tensor (+, × or grad and curl) modes in general relativity. We will denote quantities calculated using the $\hat{k}$-convention with an overbar, e.g., $\bar{R}^{A}(f,\hat{k})$.

G.1 General relationship between the response functions

Plane wave expansion:

$$h_{ab}(t,\vec{x}) = \int_{-\infty}^{\infty} \int d^2\Omega_{\hat{n}}\, h_{ab}(f,\hat{n}) e^{i2\pi f(t+\hat{n}\cdot\vec{x}/c)} \,. \tag{G.1}$$

Detector response:

$$\begin{aligned} h(t) &= \int_{-\infty}^{\infty} d\tau \int d^3y\, R^{ab}(\tau,\vec{y}) h_{ab}(t-\tau,\vec{x}-\vec{y}) \\ &= \int_{-\infty}^{\infty} df \int d^2\Omega_{\hat{n}}\, R^{ab}(f,\hat{n}) h_{ab}(f,\hat{n}) e^{i2\pi f t} \,, \end{aligned} \tag{G.2}$$

where

$$R^{ab}(f,\hat{n}) = e^{i2\pi f \hat{n}\cdot\vec{x}/c} \int_{-\infty}^{\infty} d\tau \int d^3y\, R^{ab}(\tau,\vec{y}) e^{-i2\pi f(\tau+\hat{n}\cdot\vec{y}/c)} \,. \tag{G.3}$$

Note that compared to an expansion in terms of the direction of propagation $\hat{k}$, we have:

$$R^{ab}(f,\hat{n}) = \bar{R}^{ab}(f,\hat{k})\big|_{\hat{k}=-\hat{n}} \,. \tag{G.4}$$

This is the general relationship between the response functions for the two approaches.

G.2 Polarization basis response functions

The response functions in the polarization basis are given by:

$$R^A(f,\hat{n}) = R^{ab}(f,\hat{n})e^A_{ab}(\hat{n})\,, \tag{G.5}$$

where $A = \{+, \times, X, Y, B, L\}$ label the tensor, vector, and scalar polarization modes (two for each). Since the polarization basis tensors $e^A_{ab}(\hat{n})$ are the same for the two approaches, it follows from (G.4) that

$$R^A(f,\hat{n}) = \bar{R}^{ab}(f,\hat{k})e^A_{ab}(\hat{n})\big|_{\hat{k}=-\hat{n}}\,. \tag{G.6}$$

If we further use the transformation properties of the polarization basis tensors $e^A_{ab}(\hat{n})$ under a parity transformation (i.e., $\hat{n} \to -\hat{n}$) we have:

$$\begin{aligned}
R^+(f,\hat{n}) &= \bar{R}^+(f,\hat{k})\big|_{\hat{k}=-\hat{n}}\,,\\
R^\times(f,\hat{n}) &= -\bar{R}^\times(f,\hat{k})\big|_{\hat{k}=-\hat{n}}\,,\\
R^X(f,\hat{n}) &= -\bar{R}^X(f,\hat{k})\big|_{\hat{k}=-\hat{n}}\,,\\
R^Y(f,\hat{n}) &= \bar{R}^Y(f,\hat{k})\big|_{\hat{k}=-\hat{n}}\,,\\
R^B(f,\hat{n}) &= \bar{R}^B(f,\hat{k})\big|_{\hat{k}=-\hat{n}}\,,\\
R^L(f,\hat{n}) &= \bar{R}^L(f,\hat{k})\big|_{\hat{k}=-\hat{n}}\,.
\end{aligned} \tag{G.7}$$

Note that in terms of standard angular coordinates (θ, ϕ) on the sphere, the substitution $\hat{k} = -\hat{n}$ corresponds to

$$\theta \to \pi - \theta\,, \qquad \phi \to \phi + \pi\,, \tag{G.8}$$

for which

$$\begin{aligned}
\sin\theta &\to \sin\theta\,,\\
\cos\theta &\to -\cos\theta\,,\\
\sin\phi &\to -\sin\phi\,,\\
\cos\phi &\to -\cos\phi\,.
\end{aligned} \tag{G.9}$$

G.3 Spherical harmonic basis response functions

The response functions in the spherical harmonic basis are given by:

$$R^P_{(lm)}(f) = \int d^2\Omega_{\hat{n}}\, R^{ab}(f,\hat{n})Y^P_{(lm)ab}(\hat{n})\,, \tag{G.10}$$

where $P = \{G, C, V_G, V_C, B, L\}$ label the tensor, vector, and scalar spherical harmonic modes. If we use the transformation properties of the spherical harmonics $Y^P_{(lm)ab}(\hat{n})$

under a parity transformation, it follows that:

$$
\begin{aligned}
R^{G}_{(lm)}(f) &= (-1)^{l}\,\bar{R}^{G}_{(lm)}(f)\,,\\
R^{C}_{(lm)}(f) &= (-1)^{l+1}\,\bar{R}^{C}_{(lm)}(f)\,,\\
R^{V_G}_{(lm)}(f) &= (-1)^{l}\,\bar{R}^{V_G}_{(lm)}(f)\,,\\
R^{V_C}_{(lm)}(f) &= (-1)^{l+1}\,\bar{R}^{V_C}_{(lm)}(f)\,,\\
R^{B}_{(lm)}(f) &= (-1)^{l}\,\bar{R}^{B}_{(lm)}(f)\,,\\
R^{L}_{(lm)}(f) &= (-1)^{l}\,\bar{R}^{L}_{(lm)}(f)\,.
\end{aligned}
\tag{G.11}
$$

Thus, the curl modes (both tensor and vector) involve a factor of $(-1)^{l+1}$, while all the other modes involve a factor of $(-1)^{l}$.

References

[1] Aasi, J. et al., "Search for long-lived gravitational-wave transients coincident with long gamma-ray bursts", Phys. Rev. D, **88**(12), 122004 (December 2013). [DOI], [ADS], [arXiv:1309.6160 [astro-ph.HE]].

[2] Aasi, J. et al. (VIRGO, LIGO Scientific), "Improved Upper Limits on the Stochastic Gravitational-Wave Background from 20092010 LIGO and Virgo Data", *Phys. Rev. Lett.*, **113**(23), 231101 (2014). [DOI], [arXiv:1406.4556 [gr-qc]].

[3] Aasi, J. et al., "Advanced LIGO", *Class. Quantum Grav.*, **32**, 074001 (2015). [DOI], [arXiv:1411.4547 [gr-qc]].

[4] Aasi, J. et al., "Searching for stochastic gravitational waves using data from the two colocated LIGO Hanford detectors", Phys. Rev. D, **91**(2), 022003 (January 2015). [DOI], [ADS], [arXiv:1410.6211 [gr-qc]].

[5] Abadie, J. et al. (LIGO Scientific Collaboration and Virgo Collaboration), "Directional Limits on Persistent Gravitational Waves Using LIGO S5 Science Data", *Phys. Rev. Lett.*, **107**, 271102 (Dec 2011). [DOI]URL: `http://link.aps.org/doi/10.1103/PhysRevLett.107.271102`.

[6] Abbate, Salvatore F. et al., "The Cassini gravitational wave experiment", *Proc. SPIE Int. Soc. Opt. Eng.*, **4856**, 90–97 (2003). [DOI].

[7] Abbott, B. et al., "Upper limits on a stochastic background of gravitational waves", *Physical Review Letters*, **95**, 221101 (2005).

[8] Abbott, B. et al., "Searching for a Stochastic Background of Gravitational Waves with the Laser Interferometer Gravitational-Wave Observatory", *Astrophysical Journal*, **659**, 918 (2007).

[9] Abbott, B. P. et al., "All-sky search for long-duration gravitational wave transients with initial LIGO", Phys. Rev. D, **93**(4), 042005 (February 2016). [DOI], [ADS], [arXiv:1511.04398 [gr-qc]].

[10] Abbott, B. P. et al. (Virgo, LIGO Scientific), "GW150914: Implications for the stochastic gravitational wave background from binary black holes", *Phys. Rev. Lett.*, **116**(13), 131102 (2016). [DOI], [arXiv:1602.03847 [gr-qc]].

[11] Abbott, B. P. et al., "GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence", *Physical Review Letters*, **116**(24), 241103 (June 2016). [DOI], [ADS].

[12] Abbott, B. P. et al., "Observation of Gravitational Waves from a Binary Black Hole Merger", *Physical Review Letters*, **116**(6), 061102 (February 2016). [DOI], [ADS], [arXiv:1602.03837 [gr-qc]].

[13] Abbott, B. P. et al., "Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO and Advanced Virgo", *Living Reviews in Relativity*, **19** (February 2016). [DOI], [ADS], [arXiv:1304.0670 [gr-qc]].

[14] Abramowitz, M. and Stegun, I. A., *Handbook of Mathematical Functions*, (Dover, New York, 1972).

[15] Acernese, F. et al., "Advanced Virgo: a second-generation interferometric gravitational wave detector", *Class. Quantum Grav.*, **32**(2), 024001 (2015). [DOI], [arXiv:1408.3978 [gr-qc]].

[16] Adams, Matthew R. and Cornish, Neil J., "Discriminating between a Stochastic Gravitational Wave Background and Instrument Noise", *Phys. Rev.*, **D82**, 022002 (2010). [DOI], [arXiv:1002.1291 [gr-qc]].

[17] Adams, Matthew R. and Cornish, Neil J., "Detecting a Stochastic Gravitational Wave Background in the presence of a Galactic Foreground and Instrument Noise", *Phys. Rev.*, **D89**(2), 022001 (2014). [DOI], [arXiv:1307.4116 [gr-qc]].

[18] Ade, P. A. et al. (BICEP2 Collaboration), "Detection of B-Mode Polarization at Degree Angular Scales by BICEP2", Phys. Rev. Lett., **112**, 241101 (Jun 2014).

[19] Ade, P. A. R. et al. (BICEP2), "Detection of B-Mode Polarization at Degree Angular Scales by BICEP2", *Phys. Rev. Lett.*, **112**(24), 241101 (2014). [DOI], [arXiv:1403.3985 [astro-ph.CO]].

[20] Ade, P. A. R. et al. (BICEP2, Planck), "Joint Analysis of BICEP2/$Keck?Array$ and *Planck* Data", *Phys. Rev. Lett.*, **114**, 101301 (2015). [DOI], [arXiv:1502.00612 [astro-ph.CO]].

[21] Alexander, S. H., Peskin, M. E. and Sheikh-Jabbari, M. M., "Leptogenesis from Gravity Waves in Models of Inflation", *Physical Review Letters*, **96**(8), 081301 (February 2006). [DOI], [ADS], [hep-th/0403069].

[22] Allen, B., "The Stochastic Gravity-Wave Background: Sources and Detection", in Marck, J.-A. and Lasota, J.-P., eds., *Relativistic Gravitation and Gravitational Radiation*, p. 373, (1997). [ADS], [gr-qc/9604033].

[23] Allen, B., Creighton, J. D., Flanagan, É. É. and Romano, J. D., "Robust statistics for deterministic and stochastic gravitational waves in non-Gaussian noise: Frequentist analyses", Phys. Rev. D, **65**(12), 122002 (June 2002). [DOI], [ADS], [gr-qc/0105100].

[24] Allen, Bruce, Creighton, Jolien D. E., Flanagan, Eanna E. and Romano, Joseph D., "Robust statistics for deterministic and stochastic gravitational waves in nonGaussian noise. 2. Bayesian analyses", *Phys. Rev.*, **D67**, 122002 (2003). [DOI], [arXiv:gr-qc/0205015 [gr-qc]].

[25] Allen, B. and Ottewill, A. C., "Detection of anisotropies in the gravitational-wave stochastic background", Phys. Rev. D, **56**, 545–563 (July 1997). [DOI], [ADS], [gr-qc/9607068].

[26] Allen, B. and Romano, J.D., "Detecting a stochastic background of gravitational radiation: Signal processing strategies and sensitivities", Phys. Rev. D, **59**, 102001 (1999).

[27] Anderson, T. W. and Darling, D. A., "A Test of Goodness of Fit", *Journal of the American Statistical Association*, **49**(268), 765–769 (1954). [DOI], [http://www.tandfonline.com/doi/pdf/10.1080/01621459.1954.10501232]URL: `http://www.tandfonline.com/doi/abs/10.1080/01621459.1954.10501232`.

[28] Anholm, M., Ballmer, S., Creighton, J. D. E., Price, L. R. and Siemens, X., "Optimal strategies for gravitational wave stochastic background searches in pulsar timing data", Phys. Rev. D, **79**(8), 084030 (April 2009). [DOI], [ADS], [arXiv:0809.0701 [gr-qc]].

[29] Armano, M. et al., "Sub-Femto-g Free Fall for Space-Based Gravitational Wave Observatories: LISA Pathfinder Results", *Physical Review Letters*, **116**(23), 231101 (June 2016). [DOI], [ADS].

[30] Armstrong, John W., "Low-Frequency Gravitational Wave Searches Using Spacecraft Doppler Tracking", *Living Reviews in Relativity*, **9**(1) (2006). [DOI]URL: `http://www.livingreviews.org/lrr-2006-1`.

[31] Arzoumanian, Z. et al., "The NANOGrav Nine-year Data Set: Limits on the Isotropic Stochastic Gravitational Wave Background", ApJ, **821**, 13 (April 2016). [DOI], [ADS], [arXiv:1508.03024].

[32] Arzoumanian, Z. et al. (NANOGrav), "The NANOGrav Nine-year Data Set: Limits on the Isotropic Stochastic Gravitational Wave Background", *Astrophys. J.*, **821**(1), 13 (2016). [DOI], [arXiv:1508.03024 [astro-ph.GA]].

[33] Aso, Y., Michimura, Y., Somiya, K., Ando, M., Miyakawa, O., Sekiguchi, T., Tatsumi, D. and Yamamoto, H., "Interferometer design of the KAGRA gravitational wave detector", Phys. Rev. D, **88**(4), 043007 (August 2013). [DOI], [ADS], [arXiv:1306.6747 [gr-qc]].

[34] Ballmer, S. W., "A radiometer for stochastic gravitational waves", *Classical and Quantum Gravity*, **23**, S179–S185 (April 2006). [DOI], [ADS], [gr-qc/0510096].

[35] Ballmer, S. W., *LIGO interferometer operating at design sensitivity with application to gravitational radiometry*, Ph.D. thesis, (Massachusetts Institute of Technology, Massachusetts, USA, 2006). [ADS].

[36] Barack, Leor and Cutler, Curt, "Confusion noise from LISA capture sources", *Phys. Rev.*, **D70**, 122002 (2004). [DOI], [arXiv:gr-qc/0409010 [gr-qc]].

[37] Bender, P. L. and Hils, D., "Confusion noise level due to galactic and extragalactic binaries", *Classical and Quantum Gravity*, **14**, 1439–1444 (June 1997). [DOI], [ADS].

[38] Bennett, C. L. et al. (WMAP), "First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Preliminary maps and basic results", *Astrophys. J. Suppl.*, **148**, 1–27 (2003). [DOI], [arXiv:astro-ph/0302207 [astro-ph]].

[39] BICEP2/Keck and Planck Collaborations et al., "Joint Analysis of BICEP2/Keck Array and Planck Data", *Physical Review Letters*, **114**(10), 101301 (March 2015). [DOI], [ADS], [arXiv:1502.00612].

[40] Boas, Mary L., *Mathematical Methods in the Physical Sciences, Third Edition*, (John Wiley & Sons, Inc., United States of America, 2006).

[41] Chamberlin, S. J. and Siemens, X., "Stochastic backgrounds in alternative theories of gravity: Overlap reduction functions for pulsar timing arrays", Phys. Rev. D, **85**(8), 082001 (April 2012). [DOI], [ADS], [arXiv:1111.5661 [astro-ph.HE]].

[42] Christensen, N., *On Measuring the Stochastic Gravitational Radiation Background with Laser Interferometric Antennas*, Ph.D. thesis, (Massachusetts Institute of Technology, 1990).

[43] Christensen, N., "Measuring the stochastic gravitational-radiation background with laser-interferometric antennas", Phys. Rev. D, **46**, 5250–5266 (December 1992). [DOI], [ADS].

[44] Cornish, N.J. and Romano, J.D., "When is a gravitational-wave signal stochastic?", *Phys. Rev. D*, **92**, 042001 (2015). [gr-qc/1505.08084].

[45] Cornish, N. J., "Mapping the gravitational-wave background", *Classical and Quantum Gravity*, **18**, 4277–4291 (October 2001). [DOI], [ADS], [astro-ph/0105374].

[46] Cornish, Neil J., "Detecting a stochastic gravitational wave background with the Laser Interferometer Space Antenna", *Phys. Rev.*, **D65**, 022004 (2002). [DOI], [arXiv:gr-qc/0106058 [gr-qc]].

[47] Cornish, N. J. and Larson, S. L., "Space missions to detect the cosmic gravitational-wave background", *Classical and Quantum Gravity*, **18**, 3473–3495 (September 2001). [DOI], [ADS], [gr-qc/0103075].

[48] Cornish, Neil J. and Littenberg, Tyson B., "BayesWave: Bayesian Inference for Gravitational Wave Bursts and Instrument Glitches", *Class. Quant. Grav.*, **32**(13), 135012 (2015). [DOI], [arXiv:1410.3835 [gr-qc]].

[49] Cornish, N. J. and Romano, J. D., "Towards a unified treatment of gravitational-wave data analysis", Phys. Rev. D, **87**(12), 122003 (June 2013). [DOI], [ADS], [arXiv:1305.2934 [gr-qc]].

[50] Cornish, N. J. and Sampson, L., "Towards robust gravitational wave detection with pulsar timing arrays", Phys. Rev. D, **93**(10), 104047 (May 2016). [DOI], [ADS], [arXiv:1512.06829 [gr-qc]].

[51] Cornish, N. J. and van Haasteren, R., "Mapping the nano-Hertz gravitational wave sky", *ArXiv e-prints* (June 2014). [ADS], [arXiv:1406.4511 [gr-qc]].

[52] Coughlin, Michael W. et al., "Subtraction of correlated noise in global networks of gravitational-wave interferometers" (2016). [arXiv:1606.01011 [gr-qc]].

[53] Creighton, T., Jenet, F. A. and Price, R. H., "Pulsar Timing and Spacetime Curvature", ApJ, **693**, 1113–1117 (March 2009). [DOI], [ADS], [arXiv:0812.3941].

[54] Cutler, C., "Angular resolution of the LISA gravitational wave detector", Phys. Rev. D, **57**, 7089–7102 (June 1998). [DOI], [ADS], [gr-qc/9703068].

[55] Dahlhaus, R., "Locally Stationary Processes", *ArXiv e-prints* (September 2011). [ADS], [arXiv:1109.4174 [math.ST]].

[56] del Castillo, G. F. Torres, *3-D Spinors, Spin-Weighted Functions and their Applications*, (Springer, New York, 2003).

[57] Detweiler, S., "Pulsar timing measurements and the search for gravitational waves", ApJ, **234**, 1100–1104 (December 1979). [DOI], [ADS].

[58] Drasco, S. and Flanagan, É. É., "Detection methods for non-Gaussian gravitational wave stochastic backgrounds", Phys. Rev. D, **67**(8), 082003 (April 2003). [DOI], [ADS], [gr-qc/0210032].

[59] Dwivedi, Y. and Subba Rao, S., "A test for second order stationarity of a time series based on the Discrete Fourier Transform (Technical Report)", *ArXiv e-prints* (November 2009). [ADS], [arXiv:0911.4744 [stat.ME]].

[60] Edlund, J. A., Tinto, M., Królak, A. and Nelemans, G., "Simulation of the white dwarf white dwarf galactic background in the LISA data", *Classical and Quantum Gravity*, **22**, S913–S926 (September 2005). [DOI], [ADS], [gr-qc/0504026].

[61] Ellis, Justin and Cornish, Neil, "Transdimensional Bayesian approach to pulsar timing noise analysis", *Phys. Rev.*, **D93**(8), 084048 (2016). [DOI], [arXiv:1601.00650 [astro-ph.IM]].

[62] Estabrook, F. B., Tinto, Massimo and Armstrong, J. W., "Time delay analysis of LISA gravitational wave data: Elimination of spacecraft motion effects", *Phys. Rev.*, **D62**, 042002 (2000). [DOI].

[63] Estabrook, F. B. and Wahlquist, H. D., "Response of Doppler spacecraft tracking to gravitational radiation", *General Relativity and Gravitation*, **6**, 439–447 (October 1975). [DOI], [ADS].

[64] Feldman, G. J. and Cousins, R. D., "Unified approach to the classical statistical analysis of small signals", Phys. Rev. D, **57**, 3873–3889 (April 1998). [DOI], [ADS], [physics/9711021].

[65] Finn, L. S., Larson, S. L. and Romano, J. D., "Detecting a stochastic gravitational-wave background: The overlap reduction function", Phys. Rev. D, **79**(6), 062003 (March 2009). [DOI], [ADS], [arXiv:0811.3582 [gr-qc]].

[66] Flanagan, Éanna É., "Sensitivity of the Laser Interferometer Gravitational Wave Observatory to a stochastic background, and its dependence on the detector orientations", Phys. Rev. D, **48**, 2389 (1993).

[67] Flauger, R., Hill, J. C. and Spergel, D. N., "Toward an understanding of foreground emission in the BICEP2 region", J. Cosmology Astropart. Phys., **8**, 039 (August 2014). [DOI], [ADS], [arXiv:1405.7351].

[68] Fricke, Tobin, LIGO Stochastic Analysis Group, 20 Sep 2006 e-log entry, (2006).

[69] Gair, J., Romano, J. D., Taylor, S. and Mingarelli, C. M. F., "Mapping gravitational-wave backgrounds using methods from CMB analysis: Application to pulsar timing arrays", Phys. Rev. D, **90**(8), 082001 (October 2014). [DOI], [ADS], [arXiv:1406.4664 [gr-qc]].

[70] Gair, J. R., Feroz, F., Babak, S., Graff, P., Hobson, M. P., Petiteau, A. and Porter, E. K., "Nested sampling as a tool for LISA data analysis", *Journal of Physics Conference Series*, **228**(1), 012010 (May 2010). [DOI], [ADS].

[71] Gair, J. R., Romano, J. D. and Taylor, S. R., "Mapping gravitational-wave backgrounds of arbitrary polarisation using pulsar timing arrays", Phys. Rev. D, **92**(10), 102003 (November 2015). [DOI], [ADS], [arXiv:1506.08668 [gr-qc]].

[72] Gair, J. R., Vallisneri, M., Larson, S. L. and Baker, J. G., "Testing General Relativity with Low-Frequency, Space-Based Gravitational-Wave Detectors", *Living Reviews in Relativity*, **16** (September 2013). [ADS], [arXiv:1212.5575 [gr-qc]].

[73] Galloway, D. K., Premachandra, S., Steeghs, D., Marsh, T., Casares, J. and Cornelisse, R., "Precision Ephemerides for Gravitational-wave Searches. I. Sco X-1", ApJ, **781**, 14 (January 2014). [DOI], [ADS], [arXiv:1311.6246 [astro-ph.HE]].

[74] Giampieri, G. and Polnarev, A. G., "Detecting an anisotropic gravitational wave background with a space-borne interferometer", MNRAS, **291**, 149–161 (October 1997). [DOI], [ADS].

[75] Goldberg, J. N., MacFarlane, A. J., Newman, E. T., Rohrlich, F. and Sudarshan, E. C. G., "Spin-s Spherical Harmonics and $\eth$", *Journal of Mathematical Physics*, **8**(11), 2155 (1967).

[76] Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M. and Bartelmann, M., *Astrophysical Journal*, **622**, 759 (2005).

[77] Gregory, Phil C., *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica Support*, (Cambridge University Press, Cambridge, UK, New York, NY, 2005).

[78] Grishchuk, L. P., "Primordial gravitons and possibility of their observation", *Journal of Experimental and Theoretical Physics Letters*, **23**, 293 (1976).

[79] Hanson, D. et al. (SPTpol), "Detection of B-mode Polarization in the Cosmic Microwave Background with Data from the South Pole Telescope", *Phys. Rev. Lett.*, **111**(14), 141301 (2013). [DOI], [arXiv:1307.5830 [astro-ph.CO]].

[80] Hartle, James B., *Gravity: An Introduction to Einstein's General Relativity*, (Addison Wesley, 2002).

[81] Hellings, R. W., "Gravitational Wave Detection at Low and Very Low Frequencies", in Blair, D. G., ed., *The Detection of Gravitational Waves*, p. 453, (1991). [ADS].

[82] Hellings, R. W. and Downs, G. S., "Upper limits on the istotropic gravitational radiation background from pulsar timing analysis", ApJ, **265**, L39–L42 (1983).

[83] Helstrom, Carl W., *Statistical Theory of Signal Detection, 2nd edition*, (Pergamon Press, Oxford, London, Edinburgh, New York, Toronto, Sydney, Paris, Braunschweig, 1968).

[84] Hils, D. and Bender, P. L., "Gravitational Radiation from Helium Cataclysmics", ApJ, **537**, 334–341 (July 2000). [DOI], [ADS].

[85] Hils, D., Bender, P. L. and Webbink, R. F., "Gravitational radiation from the Galaxy", ApJ, **360**, 75–94 (September 1990). [DOI], [ADS].

[86] Himemoto, Yoshiaki, Taruya, Atsushi, Kudoh, Hideaki and Hiramatsu, Takashi, "Detecting a stochastic background of gravitational waves in the presence of non-Gaussian noise: A Performance of generalized cross-correlation statistic", *Phys. Rev.*, **D75**, 022003 (2007). [DOI], [arXiv:gr-qc/0607015 [gr-qc]].

[87] Hogan, C. J. and Bender, P. L., "Estimating stochastic gravitational wave backgrounds with the Sagnac calibration", Phys. Rev. D, **64**(6), 062002 (September 2001). [DOI], [ADS], [astro-ph/0104266].

[88] Howson, C. and Urbach, P., "Bayesian reasoning in science", Nature, **350**, 371–374 (April 1991). [DOI], [ADS].

[89] Howson, Colin and Urbach, Peter, *Scientific Reasoning: The Bayesian approach*, (Open Court Publishing, Peru, Illinois, 2006).

[90] Hu, W. and White, M., "CMB anisotropies: Total angular momentum method", *Physical Review D*, **56**, 596–615 (July 1997). [DOI], [ADS].

[91] Imkeller, Peter and Storch, Jin-Song von, *Stochastic Climate Models*, Progress in Probability, Volume 49, (Birkhauser, Basel, Boston, Berlin, 2001).

[92] Iyer, Bala et al., *LIGO-India*, LIGO-M1100296, (2011). `https://dcc.ligo.org/LIGO-M1100296/public/main`.

[93] Jackson, J. D., *Classical Electrodynamics, 3rd Edition*, (1998). [ADS].

[94] Jaffe, A. H. and Backer, D. C., "Gravitational Waves Probe the Coalescence Rate of Massive Black Hole Binaries", ApJ, **583**, 616–631 (February 2003). [DOI], [ADS], [astro-ph/0210148].

[95] Janssen, Gemma et al., "Gravitational wave astronomy with the SKA", *PoS*, **AASKA14**, 037 (2015). [arXiv:1501.00127 [astro-ph.IM]].

[96] Jaynes, Edwin T., *Probability Theory: Logic of Science*, (Cambridge University Press, Cambridge, 2003).

[97] Jenet, F. A. and Romano, J. D., "Understanding the gravitational-wave Hellings and Downs curve for pulsar timing arrays in terms of sound and electromagnetic waves", *American Journal of Physics*, **83**, 635–645 (July 2015). [DOI], [ADS], [arXiv:1412.1142 [gr-qc]].

[98] Kamionkowski, M., Kosowsky, A. and Stebbins, A., "Statistics of cosmic microwave background polarization", Phys. Rev. D, **55**, 7368–7388 (1997).

[99] Kamionkowski, M. and Kovetz, E. D., "The Quest for B Modes from Inflationary Gravitational Waves", *ArXiv e-prints* (October 2015). [ADS], [arXiv:1510.06042].

[100] Kass, Robert E. and Raftery, Adrian E., "Bayes Factors", *Journal of the American Statistical Association*, **90**(430), 773–795 (June 1995). [DOI].

[101] Kato, R. and Soda, J., "Probing circular polarization in stochastic gravitational wave background with pulsar timing arrays", Phys. Rev. D, **93**(6), 062003 (March 2016). [DOI], [ADS], [arXiv:1512.09139 [gr-qc]].

[102] Knox, Lloyd and Turner, Michael S., "Detectability of tensor perturbations through CBR anisotropy", *Phys. Rev. Lett.*, **73**, 3347–3350 (1994). [DOI], [arXiv:astro-ph/9407037 [astro-ph]].

[103] Koop, M. J. and Finn, L. S., "Physical response of light-time gravitational wave detectors", Phys. Rev. D, **90**(6), 062002 (September 2014). [DOI], [ADS], [arXiv:1310.2871 [gr-qc]].

[104] Kruschke, John K, "Bayesian estimation supersedes the t test.", *Journal of Experimental Psychology: General*, **142**(2), 573–603 (2013).

[105] Kudoh, H. and Taruya, A., "Probing anisotropies of gravitational-wave backgrounds with a space-based interferometer: Geometric properties of antenna patterns and their angular power", Phys. Rev. D, **71**(2), 024025 (January 2005). [DOI], [ADS], [gr-qc/0411017].

[106] Larson, Shane L., Hiscock, William A. and Hellings, Ronald W., "Sensitivity curves for spaceborne gravitational wave interferometers", *Phys. Rev.*, **D62**, 062001 (2000). [DOI], [arXiv:gr-qc/9909080 [gr-qc]].

[107] Lasky, Paul D. et al., "Gravitational-wave cosmology across 29 decades in frequency", *Phys. Rev.*, **X6**(1), 011035 (2016). [DOI], [arXiv:1511.05994 [astro-ph.CO]].

[108] Lazzarini, A. and Romano, J., "Use of Overlapping Windows in the Stochastic Background Search", *https://dcc.ligo.org/T040089/public*, **LIGO preprint** (2004).

[109] Lee, K. J., Jenet, F. A. and Price, R. H., "Pulsar Timing as a Probe of Non-Einsteinian Polarizations of Gravitational Waves", ApJ, **685**, 1304–1319 (October 2008). [DOI], [ADS].

[110] Lentati, L., Alexander, P., Hobson, M. P., Feroz, F., van Haasteren, R., Lee, K. J. and Shannon, R. M., "TEMPONEST: a Bayesian approach to pulsar timing analysis", MNRAS, **437**, 3004–3023 (January 2014). [DOI], [ADS], [arXiv:1310.2120 [astro-ph.IM]].

[111] Lentati, L. et al., "European Pulsar Timing Array Limits On An Isotropic Stochastic Gravitational-Wave Background", *Mon. Not. Roy. Astron. Soc.*, **453**(3), 2576–2598 (2015). [DOI], [arXiv:1504.03692 [astro-ph.CO]].

[112] Littenberg, Tyson B. and Cornish, Neil J., "Separating Gravitational Wave Signals from Instrument Artifacts", *Phys. Rev.*, **D82**, 103007 (2010). [DOI], [arXiv:1008.1577 [gr-qc]].

[113] Littenberg, T. B. and Cornish, N. J., "Bayesian inference for spectral estimation of gravitational wave detector noise", Phys. Rev. D, **91**(8), 084034 (April 2015). [DOI], [ADS], [arXiv:1410.3852 [gr-qc]].

[114] M. B. Priestley, T. Subba Rao, "A Test for Non-Stationarity of Time-Series", *Journal of the Royal Statistical Society. Series B (Methodological)*, **31**(1), 140–149 (1969)URL:
`http://www.jstor.org/stable/2984336`.

[115] Maggiore, M., "Gravitational wave experiments and early universe cosmology", Phys. Rep., **331**, 283–367 (July 2000). [DOI], [ADS], [gr-qc/9909001].

[116] Martellini, L. and Regimbau, T., "Semiparametric approach to the detection of non-Gaussian gravitational wave stochastic backgrounds", Phys. Rev. D, **89**(12), 124009 (June 2014). [DOI], [ADS], [arXiv:1405.5775].

[117] Messenger, C. et al., "Gravitational waves from Scorpius X-1: A comparison of search methods and prospects for detection with advanced detectors", Phys. Rev. D, **92**(2), 023006 (July 2015). [DOI], [ADS], [arXiv:1504.05889 [gr-qc]].

[118] Messiah, A., *Quantum Mechanics, Vol. 2*, (North Holland, Amsterdam, Netherlands, 1962).

[119] Michelson, P. F., "On detecting stochastic background gravitational radiation with terrestrial detectors", MNRAS, **227**, 933–941 (August 1987). [DOI], [ADS].

[120] Mingarelli, C. M. F., Sidery, T., Mandel, I. and Vecchio, A., "Characterizing gravitational wave stochastic background anisotropy with pulsar timing arrays", Phys. Rev. D, **88**(6), 062005 (September 2013). [DOI], [ADS], [arXiv:1306.5394 [astro-ph.HE]].

[121] Misner, C. W., Thorne, K. S. and Wheeler, J. A., *Gravitation*, (W.H. Freeman and Company, San Francisco, 1973).

[122] Mitra, S., Dhurandhar, S., Souradeep, T., Lazzarini, A., Mandic, V., Bose, S. and Ballmer, S., "Gravitational wave radiometry: Mapping a stochastic gravitational wave background", Phys. Rev. D, **77**(4), 042002 (February 2008). [DOI], [ADS], [arXiv:0708.2728 [gr-qc]].

[123] Moore, C. J., Cole, R. H. and Berry, C. P. L., "Gravitational-wave sensitivity curves", *Class. Quant. Grav.*, **32**(1), 015014 (2015). [DOI], [arXiv:1408.0740 [gr-qc]].

[124] Moore, Christopher J., Taylor, Stephen R. and Gair, Jonathan R., "Estimating the sensitivity of pulsar timing arrays", *Class. Quant. Grav.*, **32**(5), 055004 (2015). [DOI], [arXiv:1406.5199 [astro-ph.IM]].

[125] Mortonson, M. J. and Seljak, U., "A joint analysis of Planck and BICEP2 B modes including dust polarization uncertainty", J. Cosmology Astropart. Phys., **10**, 035 (October 2014). [DOI], [ADS], [arXiv:1405.5857].

[126] Muller, Ulrich K., "Size and power of tests of stationarity in highly autocorrelated time series", *Journal of Econometrics*, **128**(2), 195 – 213 (2005). [DOI]URL: `http://www.sciencedirect.com/science/article/pii/S030440760400154X`.

[127] Nason, G. P., Von Sachs, R. and Kroisandt, G., "Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **62**(2), 271–292 (2000). [DOI]URL: `http://dx.doi.org/10.1111/1467-9868.00231`.

[128] Nelemans, G., Yungelson, L. R. and Portegies Zwart, S. F., "The gravitational wave signal from the Galactic disk population of binaries containing two compact objects", A&A, **375**, 890–898 (September 2001). [DOI], [ADS], [astro-ph/0105221].

[129] Newman, E. T. and Penrose, R., *Journal of Mathematical Physics*, **7**, 863 (1966).

[130] Nishizawa, A., Taruya, A., Hayama, K., Kawamura, S. and Sakagami, M.-A., "Probing nontensorial polarizations of stochastic gravitational-wave backgrounds with ground-based laser interferometers", Phys. Rev. D, **79**(8), 082002 (April 2009). [DOI], [ADS], [arXiv:0903.0528 [astro-ph.CO]].

[131] Olive, K. A. et al. (Particle Data Group), "Review of Particle Physics", *Chin. Phys.*, **C38**, 090001 (2014). [DOI].

[132] Parida, A., Mitra, S. and Jhingan, S., "Component separation of a isotropic Gravitational Wave Background", J. Cosmology Astropart. Phys., **4**, 024 (April 2016). [DOI], [ADS], [arXiv:1510.07994].

[133] Penzias, A. A. and Wilson, R. W., "A Measurement of Excess Antenna Temperature at 4080 Mc/s.", ApJ, **142**, 419–421 (July 1965). [DOI], [ADS].

[134] Phinney, S. et al., ""The Big Bang Observer: Direct detection of gravitational waves from the birth of the Universe to the Present"", *NASA Mission Concept Study* (2004).

[135] Planck Collaboration et al., "Planck 2015 results. XIII. Cosmological parameters", *ArXiv e-prints* (February 2015). [ADS], [arXiv:1502.01589].

[136] Press, William H., Teukolsky, Saul A., Vetterling, William T. and Flannery, Brian P., *Numerical Recipes in C*, (Cambridge University Press, 1992).

[137] Prince, Thomas A., Tinto, Massimo, Larson, Shane L. and Armstrong, J. W., "The LISA optimal sensitivity", *Phys. Rev.*, **D66**, 122002 (2002). [DOI], [arXiv:gr-qc/0209039 [gr-qc]].

[138] Principe, Maria and Pinto, Innocenzo M., "Modeling the impulsive noise component and its effect on the operation of a simple coherent network algorithm for detecting unmodeled gravitational wave bursts", *Class. Quant. Grav.*, **25**, 075013 (2008). [DOI], [arXiv:0806.4574 [gr-qc]].

[139] Regimbau, T., "The astrophysical gravitational wave stochastic background", *Research in Astronomy and Astrophysics*, **11**, 369–390 (April 2011). [DOI], [ADS], [arXiv:1101.2762].

[140] Regimbau, T. and Hughes, Scott A., "Gravitational-wave confusion background from cosmological compact binaries: Implications for future terrestrial detectors", *Phys. Rev. D*, **79**, 062002 (Mar 2009). [DOI].

[141] Regimbau, Tania and Mandic, Vuk, "Astrophysical Sources of Stochastic Gravitational-Wave Background", Class. Quantum Grav., **25**, 184018 (2008). [arXiv:astro-ph/0806.2794].

[142] Romano, J. D., Taylor, S. R., Cornish, N. J., Gair, J., Mingarelli, C. M. F. and van Haasteren, R., "Phase-coherent mapping of gravitational-wave backgrounds using ground-based laser interferometers", Phys. Rev. D, **92**(4), 042003 (August 2015). [DOI], [ADS], [arXiv:1505.07179 [gr-qc]].

[143] Rosado, Pablo, "Gravitational wave background from binary systems", Phys. Rev. D, **84**, 084004 (2011). [arXiv:gr-qc/1106.5795].

[144] Rosado, Pablo A., Sesana, Alberto and Gair, Jonathan, "Expected properties of the first gravitational wave signal detected with pulsar timing arrays", *Mon. Not. Roy. Astron. Soc.*, **451**(3), 2417–2433 (2015). [DOI], [arXiv:1503.04803 [astro-ph.HE]].

[145] Rover, Christian, “A Student-t based filter for robust signal detection”, *Phys. Rev.*, **D84**, 122004 (2011). [DOI], [arXiv:1109.0442 [physics.data-an]].

[146] Röver, C., Messenger, C. and Prix, R., “Bayesian versus frequentist upper limits”, *ArXiv e-prints* (March 2011). [ADS], [arXiv:1103.2987 [physics.data-an]].

[147] Sathyaprakash, B. S. and Schutz, B. F., “Physics, Astrophysics and Cosmology with Gravitational Waves”, *Living Reviews in Relativity*, **12** (March 2009). [DOI], [ADS], [arXiv:0903.0338 [gr-qc]].

[148] Schutz, B. F., *A first course in general relativity*, (Cambridge University Press, Cambridge, 1985).

[149] Seier, Edith, “International Encyclopedia of Statistical Science”, Normality Tests: Power Comparison, pp. 1000–1003, (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011). [DOI]URL:
`http://dx.doi.org/10.1007/978-3-642-04898-2_421`.

[150] Sesana, A., “Systematic investigation of the expected gravitational wave signal from supermassive black hole binaries in the pulsar timing band”, *MNRAS*, **433**, L1–L5 (June 2013). [DOI], [ADS], [arXiv:1211.5375 [astro-ph.CO]].

[151] Seto, N., “Annual modulation of the galactic binary confusion noise background and LISA data analysis”, Phys. Rev. D, **69**(12), 123005 (June 2004). [DOI], [ADS], [gr-qc/0403014].

[152] Seto, N., “Non-Gaussianity Test for Discriminating Gravitational Wave Backgrounds around 0.1-1 Hz”, ApJ, **683**, L95 (August 2008). [DOI], [ADS], [arXiv:0807.1151].

[153] Seto, N., “Non-Gaussianity analysis of a gravitational wave background made by short-duration burst signals”, Phys. Rev. D, **80**(4), 043003 (August 2009). [DOI], [ADS], [arXiv:0908.0228 [gr-qc]].

[154] Seto, N. and Cooray, A., “LISA measurement of gravitational wave background anisotropy: Hexadecapole moment via a correlation analysis”, Phys. Rev. D, **70**(12), 123005 (December 2004). [DOI], [ADS], [astro-ph/0403259].

[155] Seto, N. and Taruya, A., “Measuring a Parity-Violation Signature in the Early Universe via Ground-Based Laser Interferometers”, *Physical Review Letters*, **99**(12), 121101 (September 2007). [DOI], [ADS], [arXiv:0707.0535].

[156] Seto, N. and Taruya, A., “Polarization analysis of gravitational-wave backgrounds from the correlation signals of ground-based interferometers: Measuring a circular-polarization mode”, Phys. Rev. D, **77**(10), 103001 (May 2008). [DOI], [ADS], [arXiv:0801.4185].

[157] Shannon, R. M. et al., “Gravitational waves from binary supermassive black holes missing in pulsar observations”, *Science*, **349**, 1522–1525 (September 2015). [DOI], [ADS], [arXiv:1509.07320].

[158] Shannon, R. M. et al., "Gravitational waves from binary supermassive black holes missing in pulsar observations", *Science*, **349**, 1522–1525 (September 2015). [DOI], [ADS], [arXiv:1509.07320].

[159] Siemens, X., Ellis, J., Jenet, F. and Romano, J. D., "The stochastic background: scaling laws and time to detection for pulsar timing arrays", *Classical and Quantum Gravity*, **30**(22), 224015 (November 2013). [DOI], [ADS], [arXiv:1305.3196 [astro-ph.IM]].

[160] Sivia, D. S. and Skilling, J., *Probability Theory: Logic of Science*, (Oxford University Press, Oxford, New York, 2006).

[161] Skilling, J., "Nested sampling for general Bayesian computation", *Bayesian Analysis*, **1**, 833 (2006).

[162] SPIEGELHALTER, D. J., "An omnibus test for normality for small samples", *Biometrika*, **67**(2), 493–496 (1980). [DOI], [http://biomet.oxfordjournals.org/content/67/2/493.full.pdf+html]URL: `http://biomet.oxfordjournals.org/content/67/2/493.abstract`.

[163] Steeghs, D. and Casares, J., "The Mass Donor of Scorpius X-1 Revealed", ApJ, **568**, 273–278 (March 2002). [DOI], [ADS], [astro-ph/0107343].

[164] Talukder, D., Mitra, S. and Bose, S., "Multibaseline gravitational wave radiometry", Phys. Rev. D, **83**(6), 063002 (March 2011). [DOI], [ADS], [arXiv:1012.4530 [gr-qc]].

[165] Taruya, A. and Kudoh, H., "Probing anisotropies of gravitational-wave backgrounds with a space-based interferometer. II. Perturbative reconstruction of a low-frequency skymap", Phys. Rev. D, **72**(10), 104015 (November 2005). [DOI], [ADS], [gr-qc/0507114].

[166] Taylor, S. R. and Gair, J. R., "Searching for anisotropic gravitational-wave backgrounds using pulsar timing arrays", Phys. Rev. D, **88**(8), 084001 (October 2013). [DOI], [ADS], [arXiv:1306.5395 [gr-qc]].

[167] Taylor, S. R., Lentati, L., Babak, S., Brem, P., Gair, J. R., Sesana, A. and Vecchio, A., "All correlations must die: Assessing the significance of a stochastic gravitational-wave background in pulsar-timing arrays", *ArXiv e-prints* (June 2016). [ADS], [arXiv:1606.09180 [astro-ph.IM]].

[168] Taylor, S. R., Vallisneri, M., Ellis, J. A., Mingarelli, C. M. F., Lazio, T. J. W. and van Haasteren, R., "Are We There Yet? Time to Detection of Nanohertz Gravitational Waves Based on Pulsar-timing Array Limits", ApJ, **819**, L6 (March 2016). [DOI], [ADS], [arXiv:1511.05564 [astro-ph.IM]].

[169] Taylor, S. R. et al., "Limits on Anisotropy in the Nanohertz Stochastic Gravitational Wave Background", *Physical Review Letters*, **115**(4), 041101 (July 2015). [DOI], [ADS], [arXiv:1506.08817 [astro-ph.HE]].

[170] The LIGO Scientific Collaboration et al., "Binary Black Hole Mergers in the first Advanced LIGO Observing Run", *ArXiv e-prints* (June 2016). [ADS], [arXiv:1606.04856 [gr-qc]].

[171] The LIGO Scientific Collaboration et al., "Properties of the Binary Black Hole Merger GW150914", *ArXiv e-prints* (February 2016). [ADS], [arXiv:1602.03840 [gr-qc]].

[172] The NANOGrav Collaboration et al., "The NANOGrav Nine-year Data Set: Observations, Arrival Time Measurements, and Analysis of 37 Millisecond Pulsars", ApJ, **813**, 65 (November 2015). [DOI], [ADS], [arXiv:1505.07540 [astro-ph.IM]].

[173] Thomson, D. J., "Jackknifing multiple-window spectra", in *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on*, vi, pp. VI/73–VI/76 vol.6, (Apr 1994). [DOI].

[174] Thrane, E., "Measuring the non-Gaussian stochastic gravitational-wave background: A method for realistic interferometer data", Phys. Rev. D, **87**(4), 043009 (February 2013). [DOI], [ADS], [arXiv:1301.0263 [astro-ph.IM]].

[175] Thrane, E., Ballmer, S., Romano, J. D., Mitra, S., Talukder, D., Bose, S. and Mandic, V., "Probing the anisotropies of a stochastic gravitational-wave background using a network of ground-based laser interferometers", Phys. Rev. D, **80**(12), 122002 (December 2009). [DOI], [ADS], [arXiv:0910.0858 [astro-ph.IM]].

[176] Thrane, E., Christensen, N. and Schofield, R. M. S., "Correlated magnetic noise in global networks of gravitational-wave detectors: Observations and implications", Phys. Rev. D, **87**, 123009 (Jun 2013). [DOI]URL: `http://link.aps.org/doi/10.1103/PhysRevD.87.123009`.

[177] Thrane, E., Christensen, N., Schofield, R. M. S. and Effler, A., "Correlated noise in networks of gravitational-wave detectors: Subtraction and mitigation", Phys. Rev. D, **90**, 023013 (Jul 2014). [DOI]URL: `http://link.aps.org/doi/10.1103/PhysRevD.90.023013`.

[178] Thrane, Eric and Romano, Joseph D., "Sensitivity curves for searches for gravitational-wave backgrounds", *Phys. Rev. D*, **88**, 124032 (Dec 2013). [DOI]URL: `http://link.aps.org/doi/10.1103/PhysRevD.88.124032`.

[179] Thrane, E. et al., "Long gravitational-wave transients and associated detection strategies for a network of terrestrial interferometers", Phys. Rev. D, **83**(8), 083004 (April 2011). [DOI], [ADS], [arXiv:1012.2150 [astro-ph.IM]].

[180] Tinto, Massimo, Armstrong, J. W. and Estabrook, F. B., "Discriminating a gravitational wave background from instrumental noise in the LISA detector", *Phys. Rev.*, **D63**, 021101 (2001). [DOI].

[181] Tinto, M., Armstrong, J. W. and Estabrook, F. B., "Discriminating a gravitational-wave background from instrumental noise using time-delay interferometry", *Classical and Quantum Gravity*, **18**, 4081–4086 (October 2001). [DOI], [ADS].

[182] Turner, Michael S., White, Martin J. and Lidsey, James E., "Tensor perturbations in inflationary models as a probe of cosmology", *Phys. Rev.*, **D48**, 4613–4622 (1993). [DOI], [arXiv:astro-ph/9306029 [astro-ph]].

[183] Ungarelli, C. and Vecchio, A., "Studying the anisotropy of the gravitational wave stochastic background with LISA", Phys. Rev. D, **64**(12), 121501 (December 2001). [DOI], [ADS], [astro-ph/0106538].

[184] "`http://bicepkeck.org/`".

[185] "`http://sci.esa.int/cosmic-vision/`".

[186] "`http://sci.esa.int/lisa-pathfinder/`".

[187] "`http://sci.esa.int/planck/`".

[188] "`https://www.elisascience.org/`".

[189] van Haasteren, Rutger, Levin, Yuri, McDonald, Patrick and Lu, Tingting, "On measuring the gravitational-wave background using Pulsar Timing Arrays", *Mon. Not. Roy. Astron. Soc.*, **395**, 1005 (2009). [DOI], [arXiv:0809.0791 [astro-ph]].

[190] van Haasteren, R. and Vallisneri, M., "New advances in the Gaussian-process approach to pulsar-timing data analysis", Phys. Rev. D, **90**(10), 104012 (November 2014). [DOI], [ADS], [arXiv:1407.1838 [gr-qc]].

[191] van Haasteren, R. and Vallisneri, M., "Low-rank approximations for large stationary covariance matrices, as used in the Bayesian and generalized-least-squares analysis of pulsar-timing data", MNRAS, **446**, 1170–1174 (January 2015). [DOI], [ADS], [arXiv:1407.6710 [astro-ph.IM]].

[192] van Putten, M. H., "Proposed Source of Gravitational Radiation from a Torus around a Black Hole", *Physical Review Letters*, **87**(9), 091101 (August 2001). [DOI], [ADS], [astro-ph/0107007].

[193] van Putten, M. H. P. M., "Gravitational Waveforms of Kerr Black Holes Interacting with High-Density Matter", ApJ, **684**, L91 (September 2008). [DOI], [ADS].

[194] van Putten, M. H. P. M. and Levinson, A., "Theory and Astrophysical Consequences of a Magnetized Torus around a Rapidly Rotating Black Hole", ApJ, **584**, 937–953 (February 2003). [DOI], [ADS], [astro-ph/0212297].

[195] Wainstein, L. A. and Zubakov, V. D., *Extractions of Signals from Noise*, (Dover Publications Inc., 1971).

[196] Weinberg, S., *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*, (1972). [ADS].

[197] Wigner, E. P., *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra, expanded and improved ed.*, (Academic Press, New York, 1959).