Bayesian bounds on parameter estimation accuracy for compact coalescing binary gravitational wave signals

David Nicholson and Alberto Vecchio*
Department of Physics and Astronomy, University of Wales College of Cardiff, P.O. Box 913, Cardiff CF2 3YB, United Kingdom
(Received 22 May 1997; published 11 March 1998)

A global network of very sensitive large-scale laser interferometric gravitational wave detectors is projected to be in operation by around the turn of the century. The network is anticipated to bring a range of new astrophysical information — relating to neutron stars, black holes, and the very early universe — and also new fundamental physics information, relating to the nature of gravity in the strongly nonlinear regime for example. This information is borne by gravitational waves that will typically be very much weaker than the level of intrinsic strain noise in the detectors. Sophisticated signal extraction methods will therefore be required to analyze the network's data. Here, the noisy output of a *single* laser interferometric detector is examined. A gravitational wave is assumed to have been detected in the data. This paper is concerned only with the subsequent problem of parameter estimation. Specifically, we investigate theoretical lower bounds on the minimum mean-square errors (MSE) associated with measuring the parameters that characterize the waveform. The pre-merger inspiral waveform generated by an orbiting system of neutron stars or black holes is ideal for this study. Monte Carlo measurements of the parameters of noisy inspiral waveforms have been performed elsewhere, and the results must now confront statistical signal processing theory. Three theoretical lower bounds on parameter estimation accuracy are considered here: the Cramer-Rao bound (CRB); the Weiss-Weinstein bound (WWB); and the Ziv-Zakai bound (ZZB). The CRB is the simplest and most well-known of these bounds, but suffers from a number of limitations. It has been applied a number of times already to bound gravitational wave measurement errors. The WWB and ZZB on the other hand are computationally less simple, and we apply them here to gravitational wave parameter estimation for the very first time. The CRB is known as a local bound because it assumes that the parameters one seeks to estimate are deterministic, and provides bounds on their MSE for every possible set of intrinsic parameter values. The WWB and ZZB are known as global (Bayesian) bounds because they assume that the parameter set is random, of known prior distribution. They bound the global MSE averaged over this prior distribution. We first set up a model problem in order to develop intuition about the conditions under which global bounds are more appropriate than their local counterparts. Then we obtain the WWB and ZZB for the Newtonian-form of the coalescing binary waveform, and compare them with published analytic CRB and numerical Monte Carlo results. At large signal-to-noise ratio (SNR), we find that the theoretical bounds are all identical and are attained by the Monte Carlo results. As SNR gradually drops below ~ 10, the WWB and ZZB are both found to provide increasingly tighter lower bounds than the CRB. However, at these levels of moderate SNR, there is a significant departure between all the bounds and the numerical Monte Carlo results. We argue that the WWB and ZZB are probably within a few percent of the theoretical minimum MSE attainable for this problem. The implication is that the maximum likelihood method of parameter estimation used by the Monte Carlo simulations is not the optimal estimator for this problem at low-to-moderate SNR. In fact, it is well-known that the optimal parameter estimator is the conditional mean estimator. This, unfortunately, is notoriously difficult to compute in general. We therefore advance a strategy for implementing this method efficiently, as a post-processor to the maximum likelihood estimator, in order to achieve improved accuracy in parameter estimation. [S0556-2821(98)02806-9]

PACS number(s): 04.80.Nn, 04.30.Db, 97.80.−d,

I. INTRODUCTION

A network of very sensitive instruments is presently being assembled across the globe with a view to directly detecting cosmic gravitational radiation on Earth. Although detection is the initial goal of the network, it must ultimately function as an astrophysical observatory. In this capacity it will study a range of sources of gravitational waves, such as neutron stars, black holes, and the early universe. The network will also provide data through which to learn new and fundamental physics about the gravitational field. Arguably, broadband instruments based upon the method of laser interferometry offer the best long-term potential for making astrophysical observations. Several such instruments have been funded and are being constructed at the present time. They include: LIGO, the U.S. 4 km arm-length Laser Interferometric Gravitational Wave Observatory [1]; VIRGO, a French-Italian project to construct a 3 km arm-length interferometer [2]; GEO600, a U.K.-German effort to build a 600 m arm-length interferometer [3]; TAMA, a Japanese project for the construction of a 300 m arm-length interferometer [4]. Joint observations between these interferometers are scheduled for

*Present address: Max Planck Institut für Gravitationsphysik, Albert Einstein Institut, Schlaazweg 1, D-14473 Potsdam.

soon after the turn of the century. All of the instruments will continue to improve their sensitivity incrementally for many years thereafter.

Interferometers are intrinsically noisy instruments and the presence of noise will mask the identity of all but the very strongest incident gravitational waves. This requires a very careful and sophisticated processing of the data in order to extract the valuable information that is borne by the waves [39]. Let us consider the data acquired by only a *single* detector in the network. Suppose that the detection of a gravitational wave of some assumed known form has already been made in the instrument's noisy output. Our focus of attention here is the subsequent measurement of the one or more parameters that characterize the waveform. Specifically, we would like to compute the theoretical minimum mean-square errors with which the parameters can in principle be measured. This will clearly have an impact on the astrophysical inferences that can then be drawn as a result of the observation.

The particular gravitational waveform that we will focus on in this paper is the *chirp* of gravitational radiation that precedes the coalescence of a compact binary system comprising of neutron stars (NS's) and/or black holes (BH's). Coalescing binaries are the most promising sources of gravitational waves in the long run for the LIGO-VIRGO-GEO600-TAMA detectors [5–7]. As radiation reaction drives the stars through a slow inspiral phase just prior to coalescence, the binary generates a very clean gravitational wave signal that is amenable to theoretical modeling (see [5,7–10] and references therein). LIGO and VIRGO anticipate observing the last few minutes of neutron-star–neutron-star (NS-NS) inspiral, during which the gravitational waves oscillate through $\sim 16\,000$ cycles as their frequency sweeps through the visibility bands of the detectors. The coalescing binary event rate predictions are subject to gross uncertainties. However, it is not unreasonable that there may be a few NS-NS, NS-BH, and BH-BH mergers out to a distance of ~ 200 Mpc in a period of a year [11–15].

Inspiraling binary gravitational waves are encoded with a rich suite of physical and astrophysical information. This ranges from tests of general relativity [7,16,17], through measurements of neutron star and black hole mass and spin [6,7,18–20], to new and independent inferences about the value of the cosmological parameters [21–24]. The informational content of binaries has driven gravitational wave theorists to focus much of their collective effort on waveform calculations for inspiraling binaries. In tandem with this, there has been considerable study of algorithms for analyzing noisy gravitational wave data to extract the waveform information.

The main goal of this paper is to reassess the issue of information extraction with respect to observations of coalescing binaries by an interferometric gravitational wave detector. Our motivation is the result of a recent confrontation between the theoretical Cramer-Rao low bound (CRB) on parameter estimation errors [25,26] for coalescing binary waveforms, with real measurement errors based upon application of the *maximum likelihood* (ML) method of parameter estimation to simulated data sets [27]. The ML errors were found to depart significantly from the CRB at moderate signal-to-noise ratios (SNR's) of around ~ 8. This implies one of the following: (i) The CRB is a *weak* lower bound at this SNR, in which case a tighter theoretical bound would be desirable; (ii) the CRB is a tight bound and the ML method is not the optimal estimator, in which case a more refined parameter estimation method would be desirable; or (iii) the CRB is both weak *and* the ML method is not optimal, in which case it may be desirable to seek an improved theoretical lower bound and an improved estimator.

In an attempt to discriminate between these options, we apply new theoretical bounds on the measurement accuracy to the problem. Specifically, we investigate the Weiss-Weinstein bound [28] and the Ziv-Zakai bound [29]. These are *Bayesian bounds* that are much more versatile than the more familiar CRB. Although a little more difficult to compute than the CRB, they can often be considerably tighter. This has been demonstrated for a range of parameter estimation problems in radar and sonar [30].

The paper is organized as follows. Section II is an overview of parameter estimation, emphasizing the differences between *local bounds* on parameter estimation accuracy of which the CRB is an example, and Bayesian bounds of which the Weiss-Weinstein and Ziv-Zakai bounds are examples. The Weiss-Weinstein bound is described in some detail in Sec. III. This is followed in Sec. IV by a detailed description of the Ziv-Zakai bound. In Sec. V, some of the computational issues posed by these bounds are presented. The bounds are applied, in Sec. VI, to a simple problem that illustrates the general superiority of Bayesian bounds over their local counterparts. Then, in Sec. VII, the bounds are computed for a Newtonian coalescing binary waveform immersed in Gaussian random noise of spectral density characteristic of the first-stage LIGO detectors. We compare the bounds with actual parameter estimation errors that were obtained recently from a Monte Carlo experiment designed to test the maximum-likelihood method. The main results are discussed in Sec. VIII and some pointers for future work are given.

II. SUMMARY OF PARAMETER ESTIMATION

A common problem in many fields, ranging from radar and sonar through to geophysics and astronomy, is to seek estimates for the set of parameters characterizing a waveform that is corrupted by additive Gaussian noise. Consider the observation

$$x(t) = s(t;\boldsymbol{\theta}) + n(t), \quad |t| < T/2, \tag{1}$$

and assume that the signal $s(t;\boldsymbol{\theta})$ is a known function of time for all values of the parameter vector $\boldsymbol{\theta}$, and $n(t)$ is a zero-mean Gaussian random noise process. Some measurement algorithm, which we do not need to specify in detail yet, is applied to $s(t)$ in order to extract $\boldsymbol{\theta}$. The algorithm produces an estimate $\hat{\boldsymbol{\theta}}$, with associated error $\boldsymbol{\epsilon} \equiv \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$. A statistical summary of the performance of the algorithm is contained in the error covariance matrix, given by $\mathcal{R} \equiv \langle \boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{T}} \rangle$, where $\langle \cdot \rangle$ denotes expectation. The diagonal elements of $\mathcal{R}$ are the mean-square errors (MSE's) on each individual parameter, while the off-diagonal elements represent their cross-covariances. In order to benchmark the performance of any practical parameter estimator, one would like to know the

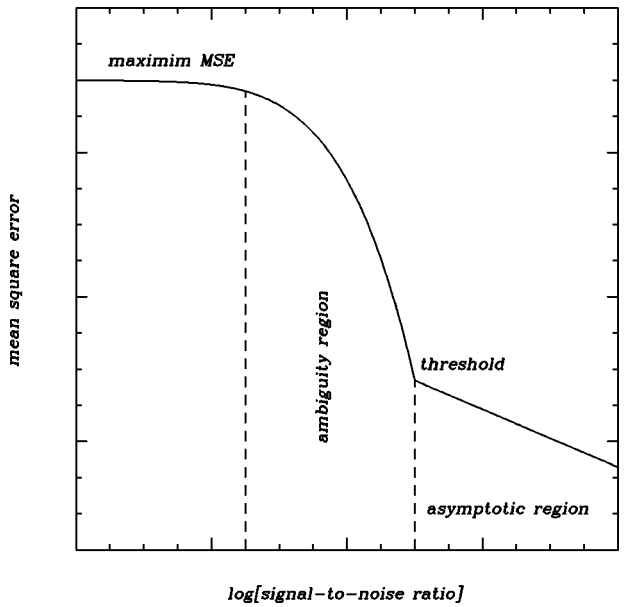

FIG. 1. Schematic representation of the behavior of the mean square error as a function of the signal-to-noise ratio in a typical nonlinear estimation problem.

theoretical *minimum* $\mathcal{R}$ for the problem at hand. If the signal parameters enter nonlinearly, then a closed-form expression for this cannot be found, and in general it is prohibitively difficult to compute numerically. A schematic picture of how the MSE will depend on the SNR in a nonlinear parameter estimation problem is given in Fig. 1. In the small error or asymptotic region, characterized by a high SNR, estimation errors are small. In the ambiguity region, where the SNR is moderate, large errors occur. When the SNR is very small, the observations provide little information and the MSE is close to that obtained simply from prior knowledge about the problem. In this paper we will be concerned with bounds that are able to characterize performance in the asymptotic and ambiguity regions. These bounds generally fall into one of two classes: local bounds or global Bayesian bounds. We will describe their main features in the next section.

A. Local bounds

The formulation of local bounds is based on the premise that the unknown parameters one seeks to measure are deterministic quantities. The bounds are local in the sense that they are placed on the MSE's for each different possible value of the intrinsic parameter vector. Local bounds have two serious limitations. First, they are restricted in application to estimators that are *unbiased*. In practice, biased estimation is often unavoidable. If the space of a parameter is finite, for example, then an unbiased estimator of it does not exist. Second, local bounds are unable to incorporate any prior information that one might have about the parameters. The Cramer-Rao bound is a familiar example of a local bound and it therefore has only limited utility. This bound states that for any *unbiased* estimator of a parameter vector $\boldsymbol{\theta}$, based upon noisy observations $\mathbf{x}$, the error covariance matrix must be larger than or equal to the inverse of the Fisher information at $\boldsymbol{\theta}$. Thus

$$\mathcal{R}_{ij}=\langle(\hat{\theta}_i-\theta_i)(\hat{\theta}_j-\theta_j)\rangle\geqslant\mathcal{J}_{ij}^{-1}, \tag{2}$$

where $\mathcal{J}$ is the Fisher information matrix whose elements are given by [35]

$$\mathcal{J}_{ij}=\left\langle\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln\Lambda(\mathbf{x};\boldsymbol{\theta})\right\rangle \tag{3}$$

and $\Lambda(\mathbf{x};\boldsymbol{\theta})$ is the likelihood ratio,

$$\Lambda(\mathbf{x};\boldsymbol{\theta})=\frac{p(\mathbf{x}|\boldsymbol{\theta})}{p(\mathbf{x}|0)}. \tag{4}$$

The CRB is not difficult to compute and it is widely invoked. In particular, it has been used almost exclusively to bound the measurement errors on the parameters of gravitational wave signals [18,19,31–34]. Moreover, it can be proved that the CRB is asymptotically attained by the maximum-likelihood method of parameter estimation [35]. As gravitational wave observations of coalescing binary signals will be rare, and the majority of detections will be made at only moderate SNR's, it is unlikely that the asymptotic conditions will be met in practice. Similar situations exist in the fields of radar and sonar, and here alternative bounds to the CRB have been considered. The Barankin bound, for example, is a local bound that can be much tighter than the CRB [36]. However, it is considerably more difficult to compute as it requires maximization over a number of free variables. Also, being a local bound, it still only applies to unbiased estimators and is unable to incorporate prior information about the parameters if this information is available.

B. Bayesian bounds

Rather than treating the unknown parameters as deterministic quantities, Bayesian bounds treat them as random variables with known prior distributions. These bounds are global in the sense that they bound MSE's on each of the parameters, averaged over their prior distributions. In contrast with their local counterparts, Bayesian bounds are not restricted in application to unbiased estimators. In fact, they lower bound the performance of *any* estimator. Also, unlike local bounds, they easily incorporate any prior information about the parameters. It is straightforward to form a Bayesian version of the CRB by simply replacing the conditional probability density $p(\mathbf{x}|\boldsymbol{\theta})$ with a joint probability density $p(\mathbf{x},\boldsymbol{\theta})$ using Bayes' theorem. However the Bayesian CRB is subject to a stringent regularity condition: It requires the prior probability density function of the parameters, $p(\boldsymbol{\theta})$, to be twice differentiable. In the common case of parameters that have uniform priors, this regularity condition is obviously not met.

Another example of a Bayesian bound is the conditional mean estimation bound (CMB) [26]. In fact this is not really a bound, since it can be attained by the *conditional mean estimator* (CME). This estimator achieves the minimum MSE and provides the benchmark against which the performance of other estimators should be compared. The CME of a scalar parameter θ, based upon noisy observations $\mathbf{x}$, is given by

$$\hat{\theta}(\mathbf{x})=\langle\theta|\mathbf{x}\rangle=\int_{-\infty}^{\infty}\theta\ p(\theta|\mathbf{x})\mathrm{d}\theta. \tag{5}$$

In the context of gravitational wave parameter estimation, the CME has been referred to as the Kallianpur-Striebel or nonlinear filter [37]. Unfortunately, the CMB is prohibitively difficult to compute for all but the simplest of problems. It generally requires multi-dimensional integrations to be performed numerically over the prior parameter space.

The complexity of the CMB has motivated the formulation of two further important Bayesian bounds — the Weiss-Weinstein bound (WWB) and the Ziv-Zakai bound (ZZB). These trade off some of the computational complexity of the CMB, and yet are only apparently a little less tight. In a range of applications the bounds have demonstrated this utility [30]. We now describe each of these bounds in some detail in the following two sections.

III. WEISS-WEINSTEIN BOUND

The Bayesian form of the CRB and the CMB discussed in the previous section belong to a general class of Bayesian bounds that Weiss and Weinstein were able to derive from the Schwarz inequality [28]. An outline of their derivation is given here. The WWB is a member of this general class, but it is free from the problems that limit the CRB and CMB. It therefore is of much more general utility than the latter two bounds. A version of the WWB for the case of a single parameter is obtained below. A statement of the multiple-parameter generalization of the scalar bound follows.

A. Single parameter

A lower bound on the error in estimating a scalar parameter θ, based upon noisy observations $\mathbf{x}$, is sought. Let $p(\mathbf{x},\theta)$ denote the joint probability density of $\mathbf{x}$ and θ. Weiss and Weinstein introduced a function $\psi(\mathbf{x},\theta)$ such that

$$\int_{-\infty}^{\infty} \mathrm{d}\theta\ \psi(\mathbf{x},\theta)p(\mathbf{x},\theta)=0 \quad \forall \mathbf{x}. \tag{6}$$

Since, for any real-valued measurable function $g(\mathbf{x})$,

$$\langle g(\mathbf{x})\psi(\mathbf{x},\theta)\rangle=\int_{-\infty}^{\infty}\mathrm{d}\mathbf{x}\ g(\mathbf{x})\int_{-\infty}^{\infty}\mathrm{d}\theta\ \psi(\mathbf{x},\theta)p(\mathbf{x},\theta)=0, \tag{7}$$

the condition in Eq. (6) implies that $\psi(\mathbf{x},\theta)$ is orthogonal to any transformation of the data $\mathbf{x}$. Subtracting $\langle \theta\psi(\mathbf{x},\theta)\rangle$ from both sides above and then applying Schwarz's inequality to the left side, results in the following:

$$\langle[\theta-g(\mathbf{x})]^2\rangle\geqslant\frac{\langle\theta\psi(\mathbf{x},\theta)\rangle^2}{\langle\psi^2(\mathbf{x},\theta)\rangle}. \tag{8}$$

As this inequality is valid for any $g(\mathbf{x})$, it sets a lower bound on the mean-square error in estimating θ from observation of $\mathbf{x}$. To underline this point we will replace $g(\mathbf{x})$ in the following by $\hat{\theta}(\mathbf{x})$. Note that the lower bound set by the right side of Eq. (8) is independent of the estimator [i.e. is absent of $g(\mathbf{x})$]. Recall that local bounds, such as the CRB, do not share this property of Bayesian bounds: They apply only to estimators that are unbiased.

It is simple to see that the Bayesian form of the CRB and the CMB are special cases of the general bound in Eq. (8). Consider choosing the function $\psi(\mathbf{x},\theta)$ as follows:

$$\psi(\mathbf{x},\theta)=\frac{\partial \ln\ p(\mathbf{x},\theta)}{\partial\theta}. \tag{9}$$

This choice satisfies the orthogonality condition (6) and generates the Bayesian CRB. Similarly, the selection

$$\psi(\mathbf{x},\theta)=\theta-\langle\theta|\mathbf{x}\rangle \tag{10}$$

leads to the CMB.

In their quest for a less restrictive bound than the CRB and the CMB, Weiss and Weinstein were led to consider a different choice for $\psi(\mathbf{x},\theta)$. They proposed the following:

$$\psi(\mathbf{x},\theta)=L^r(\mathbf{x};\theta+\delta,\theta)-L^{1-r}(\mathbf{x};\theta-\delta,\theta), \tag{11}$$

where r and δ are arbitrary real-valued scalars and $L(\mathbf{x};\theta_1,\theta_2)$ is the *likelihood ratio*,

$$L(\mathbf{x};\theta_1,\theta_2)=\frac{p(\mathbf{x},\theta_1)}{p(\mathbf{x},\theta_2)}. \tag{12}$$

This choice for $\psi(\mathbf{x},\theta)$ satisfies the orthogonality condition (6) for all combinations of δ and $0<r<1$. Substitution into Eq. (8) generates the WWB on the mean-square error, ϵ^2, in the estimation of θ:

$$\epsilon^2\geqslant\frac{\delta^2\exp[2\,\eta(r,\delta)]}{\exp[\,\eta(2r,\delta)]+\exp[\,\eta(2-2r,-\delta)]-2\exp[\,\eta(r,2\delta)]}, \tag{13}$$

where

$$\begin{aligned}\eta(r,\delta)&=\ln\ \langle L^r(\mathbf{x};\theta+\delta,\theta)\rangle=\ln\ \int_\Theta p^r(\theta+\delta)p^{1-r}(\theta)\\&\quad\times\left\{\int p^r(\mathbf{x}|\,\theta+\delta)p^{1-r}(\mathbf{x}|\,\theta)\mathrm{d}\mathbf{x}\right\}\mathrm{d}\theta\\&=\ln\ \int_\Theta p^r(\theta+\delta)p^{1-r}(\theta)\exp[\,\mu(r;\theta+\delta,\theta)]\mathrm{d}\theta.\end{aligned} \tag{14}$$

Several comments are pertinent here. First, note that the integration with respect to θ is performed over the region $\Theta=\{\theta:p(\theta)>0\}$ in order to avoid singularities. Second, the bound reduces to the Bayesian version of the CRB for $\delta\to0$. Third, the term $\mu(r;\theta+\delta,\theta)$ is a familiar one in information and communications theory: It is known as the semi-invariant moment generating function and used to bound the probability of error in binary hypothesis testing problems [26]. We shall meet it again in our discussion of the Ziv-Zakai bound in the next section.

In order to remove one of the degrees of freedom, the WWB is usually computed for $r=1/2$. It then reduces to

$$\epsilon^2\geqslant\frac{\delta^2\exp[2\,\eta(1/2,\delta)]}{2\{1-\exp[\,\eta(1/2,2\,\delta)]\}}. \tag{15}$$

The variable δ that enters the bound is usually referred to as a test point. The optimal value for δ is the one that generates the maximum bound. This value may be other than $\delta \to 0$, for which the WWB reduces to the CRB as we remarked earlier.

Weiss and Weinstein have shown how to generalize the single test point bound (15) to incorporate multiple test points. Consider a vector of N test points $\boldsymbol{\delta} \equiv (\delta_1, \delta_2, \ldots, \delta_N)$. The corresponding multiple-test point WWB is

$$\epsilon^2 \geqslant \mathbf{u} \mathcal{Q}^{-1} \mathbf{u}^{\mathsf{T}}, \tag{16}$$

where the elements of the vector $\mathbf{u}$ are

$$u_i = \delta_i, \tag{17}$$

and the elements of the matrix $\mathcal{Q}$ are

$$\mathcal{Q}_{ij} = 2 \frac{\exp[\eta(1/2, \delta_i - \delta_j)] - \exp[\eta(1/2, \delta_i + \delta_j)]}{\exp[2\eta(1/2, \delta_i)]}. \tag{18}$$

In order to evaluate Eq. (16), a matrix of dimension equal to the number of test points has to be inverted numerically. This imposes a practical restriction on the number of test points. However, as we shall see later, the WWB fortunately appears to converge quickly with increasing N.

B. Multiple parameters

Since the coalescing binary waveform is characterized by more than one parameter, we shall require a multiple parameter version of the WWB. Consider a vector of M parameters, $\boldsymbol{\theta} \equiv (\theta_i, \ldots, \theta_M)$. The WWB on the error covariance matrix $\mathcal{R}$ is obtained in a similar fashion to the single parameter bound. The result is

$$\mathcal{R} \geqslant \mathcal{H} \mathcal{G}^{-1} \mathcal{H}^{\mathsf{T}}. \tag{19}$$

The elements of the matrix $\mathcal{H}$ are the $M \times N$ test points in the multi-dimensional parameter space. The $N \times N$ matrix $\mathcal{G}$ has elements given by

$$\mathcal{G}_{ij} = 2 \frac{\exp[\eta(1/2, \boldsymbol{\delta}_i - \boldsymbol{\delta}_j)] - \exp[\eta(1/2, \boldsymbol{\delta}_i + \boldsymbol{\delta}_j)]}{\exp[\eta(1/2, \boldsymbol{\delta}_i)] \exp[\eta(1/2, \boldsymbol{\delta}_j)]}, \tag{20}$$

where $\boldsymbol{\delta}_i$ is the ith test point in the parameter space and

$$\eta(1/2, \boldsymbol{\delta}_i \boldsymbol{\delta}_j) = \ln \int_\Theta \sqrt{p(\boldsymbol{\theta} + \boldsymbol{\delta}_i) p(\boldsymbol{\theta} + \boldsymbol{\delta}_j)} \times \{ \sqrt{p(\mathbf{x} | \boldsymbol{\theta} + \boldsymbol{\delta}_i) p(\mathbf{x} | \boldsymbol{\theta} + \boldsymbol{\delta}_j)} \mathrm{d}\mathbf{x} \} \mathrm{d}\boldsymbol{\theta}. \tag{21}$$

Again, several comments are pertinent. First, integration with respect to $\boldsymbol{\theta}$ is over the region $\Theta = \{\boldsymbol{\theta} : p(\boldsymbol{\theta}) > 0\}$. Second, for a non-singular bound there must be at least M linearly independent test points. Finally, Eq. (19) reduces to the Bayesian CRB upon setting $\mathcal{H} = \delta \mathcal{I}$, where $\mathcal{I}$ is the identity matrix and the scalar $\delta \to 0$. We will turn to the practical issues involved in the computation of the scalar and vector parameter versions of the WWB in Sec. V.

IV. ZIV-ZAKAI BOUND

The theoretical foundation of the Ziv-Zakai bound is somewhat different than the WWB [29]: There do not appear to be any formal theoretical links. In common with the WWB however, the ZZB places a fundamental lower bound on the performance of *any* parameter estimator. We present a simple derivation here for the case of a single parameter with a uniform prior distribution. The multiple-parameter extension of the bound for arbitrary priors is also presented.

A. Single parameter

As a concrete example, suppose that an estimate of the difference between the arrival time of a gravitational wave at two separated detectors is required. Let us denote this parameter by θ. Now ask what is the probability of making a correct decision between two possible values, ϕ and $\phi + \Delta$, of this parameter. The likelihood ratio test (LRT) is the optimal decision scheme that produces the minimum probability of error. Instead of the LRT, consider a simpler suboptimal decision scheme in which a decision is made in favor of the ''nearest neighbor'' to some arbitrary estimate, $\hat{\theta}$, of θ. Thus,

$$\text{decide } H_0 : \theta = \phi \quad \text{if } \hat{\theta} \leqslant \phi + \frac{\Delta}{2},$$

$$\text{decide } H_1 : \theta = \phi + \Delta \quad \text{if } \hat{\theta} > \phi + \frac{\Delta}{2}. \tag{22}$$

If the two hypothesized delays are equally likely to occur, which is physically most reasonable, then the suboptimal decision scheme has a probability of error given by

$$P(\phi, \phi + \Delta) = \frac{1}{2} P\left(\hat{\theta} > \phi + \frac{\Delta}{2} \middle| \theta = \phi \right) + \frac{1}{2} P\left(\hat{\theta} \leqslant \phi + \frac{\Delta}{2} \middle| \theta = \phi + \Delta \right). \tag{23}$$

Clearly if $P_{\min}(\phi, \phi + \Delta)$ is the minimum probability of error, associated with the LRT, then

$$P_{\min}(\phi, \phi + \Delta) \leqslant \frac{1}{2} P\left(\epsilon > \frac{\Delta}{2} \middle| \phi \right) + \frac{1}{2} P\left(\epsilon \leqslant -\frac{\Delta}{2} \middle| \phi + \Delta \right), \tag{24}$$

where $\epsilon = \hat{\theta} - \theta$ denotes the estimation error. Now, suppose that θ is uniformly distributed on $[-T, T]$. In this specific example, T would represent the gravitational wave travel time between the two detectors. The inequality (24) holds good for any ϕ and Δ, in particular combinations of ϕ and Δ such that $\phi, \phi + \Delta \in [-T, T]$, or

$$-T \leqslant \phi \leqslant T - \Delta, \quad 0 < \Delta < 2T. \tag{25}$$

Integrating Eq. (24) with respect to ϕ over $[-T, T - \Delta]$ gives

$$\int_{-T}^{T-\Delta} P_{\min}(\phi,\phi+\Delta)\,\mathrm{d}\phi \leqslant \frac{1}{2}\int_{-T}^{T} P\left(|\epsilon| \geqslant \frac{\Delta}{2}\middle|\,\phi\right)\,\mathrm{d}\phi, \tag{26}$$

which can equivalently be expressed as

$$\int_{-T}^{T-\Delta} P_{\min}(\phi,\phi+\Delta)\,\mathrm{d}\phi \leqslant TH\left(\frac{\Delta}{2}\right), \tag{27}$$

where

$$H(\Delta) \equiv \frac{1}{2T}\int_{-T}^{T} P(|\epsilon| \geqslant \Delta|\ \phi)\ \mathrm{d}\phi. \tag{28}$$

Note that Eq. (27) is only useful for $\Delta \leqslant 2T$, since for $\Delta > 2T$ the integral is negative and therefore zero is a better bound. The next step is to multiply both sides of Eq. (27) by Δ/T and integrate with respect to Δ over $[0,2T]$. Noting that

$$\epsilon^2 = -\int_0^{2T} \Delta^2 \mathrm{d}\{H(\Delta)\} \tag{29}$$

is the mean-square error in the estimation of θ when the latter has a uniform prior distribution in $[-T,T]$, the integration yields

$$\epsilon^2 \geqslant \frac{1}{2T}\int_0^{2T} \Delta\ \mathrm{d}\Delta \int_{-T}^{T-\Delta} P_{\min}(\phi,\phi+\Delta)\,\mathrm{d}\Delta, \tag{30}$$

which is the ZZB in its simplest incarnation. Bellini and Tartara [38] have remarked that $H(\Delta)$ is a non-increasing function of Δ and suggested that the bound might be tightened by applying a ''valley-filling'' function to the left side of Eq. (27). Denoting this function by $V[\cdot]$, the Bellini-Tartara version of the ZZB is

$$\epsilon^2 \geqslant \frac{1}{2T}\int_0^{2T} \Delta V\left[\int_{-T}^{T-\Delta} P_{\min}(\phi,\phi+\Delta)\,\mathrm{d}\phi\right]\mathrm{d}\Delta. \tag{31}$$

The bound generalizes in a straightforward manner for an arbitrary prior, $p(\theta)$, to give

$$\epsilon^2 \geqslant \int_0^{\infty} \frac{\Delta}{2} V\left\{\int_{-\infty}^{\infty} [p(\phi)+p(\phi+\Delta)] \times P_{\min}(\phi,\phi+\Delta)\,\mathrm{d}\phi\right\}\mathrm{d}\Delta. \tag{32}$$

Although there is generally no closed form expression for $P_{\min}(\phi,\phi+\Delta)$, tight lower bounds exist [26].

B. Multiple parameters

The ZZB has only recently been extended to vector random parameters with arbitrary prior distributions [30]. Consider an M-dimensional vector random variable, $\boldsymbol{\theta}$, with prior probability distribution function (PDF) $p(\boldsymbol{\theta})$. As before, let $\hat{\boldsymbol{\theta}}$ be an estimate of $\boldsymbol{\theta}$ produced by any estimator, $\boldsymbol{\epsilon}$ the estimation error, and $\mathcal{R} = \langle \boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathsf{T}} \rangle$ the error covariance matrix. Then the following lower bound on $\mathbf{a}^{\mathsf{T}}\mathcal{R}\mathbf{a}$ for any M-dimensional vector $\mathbf{a}$ has been obtained:

$$\mathbf{a}^{\mathsf{T}}\mathcal{R}\mathbf{a} \geqslant \int_0^{\infty} \frac{\Delta}{2} V'\ \mathrm{d}\Delta, \tag{33}$$

where the valley filling function V' is now defined as

$$V' \equiv V\left\{\max \int [p(\boldsymbol{\phi})+p(\boldsymbol{\phi}+\boldsymbol{\delta})] P_{\min}(\boldsymbol{\phi},\boldsymbol{\phi}+\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\phi}\right\}, \tag{34}$$

with the maximum referred to $\boldsymbol{\delta}$. The bound is generated, as for the single parameter case, via an inequality between the probability of error in a suboptimal decision rule and the minimum probability of error associated with an LRT. However, one has now to decide between one of two possible values $\boldsymbol{\phi}$ or $\boldsymbol{\phi}+\boldsymbol{\delta}$ for the parameter vector under investigation. The suboptimal decision rule is then

$$\begin{aligned} &\text{decide } H_0 : \boldsymbol{\theta} = \boldsymbol{\phi} \quad \text{if } \mathbf{a}^{\mathsf{T}}\hat{\boldsymbol{\theta}} > \mathbf{a}^{\mathsf{T}}\boldsymbol{\phi} + \frac{\Delta}{2}, \\ &\text{decide } H_1 : \boldsymbol{\theta} = \boldsymbol{\phi}+\boldsymbol{\delta} \quad \text{if } \mathbf{a}^{\mathsf{T}}\hat{\boldsymbol{\theta}} \leqslant \mathbf{a}^{\mathsf{T}}\boldsymbol{\phi} + \frac{\Delta}{2}. \end{aligned} \tag{35}$$

The hyperplane

$$\mathbf{a}^{\mathsf{T}}\boldsymbol{\theta} = \mathbf{a}^{\mathsf{T}}\boldsymbol{\phi} + \frac{\Delta}{2}, \tag{36}$$

separating the two decision regions, passes through the midpoint of the line connecting $\boldsymbol{\phi}$ and $\boldsymbol{\phi}+\boldsymbol{\delta}$ and is perpendicular to the $\mathbf{a}$ axis. A decision is made in favor of the hypothesis that is on the same side of the separating hyperplane as the estimate $\hat{\boldsymbol{\theta}}$. The tightest bound in Eq. (33) is achieved by maximization over the vector $\boldsymbol{\delta}$, subject to the constraint $\mathbf{a}^{\mathsf{T}}\boldsymbol{\delta} = \Delta$. This constraint does not determine the vector $\boldsymbol{\delta}$ uniquely. In order to satisfy it, $\boldsymbol{\delta}$ must be composed of a fixed component along the $\mathbf{a}$ axis and an arbitrary component orthogonal to $\mathbf{a}$. That is

$$\boldsymbol{\delta} = \frac{\Delta}{\|\mathbf{a}\|^2}\mathbf{a} + \mathbf{b}, \tag{37}$$

where

$$\mathbf{a}^{\mathsf{T}}\mathbf{b} = 0, \tag{38}$$

and there are $M-1$ degrees of freedom in choosing $\boldsymbol{\delta}$ via the vector $\mathbf{b}$. Simply setting $\mathbf{b}=0$ results in hypotheses that are separated by the smallest Euclidean distance. However, this does not necessarily guarantee the largest probability of error. A maximization over $\boldsymbol{\delta}$ can improve the bound.

V. COMPUTATIONAL ISSUES

In this section the CRB, WWB, and ZZB are reduced to forms that are appropriate for calculating error bounds on the parameters of a signal that is immersed in a background of *Gaussian* random noise.

The signal waveform $s(t,\boldsymbol{\theta})$ is parametrized by a vector, $\boldsymbol{\theta}$, for which an estimate is sought. Noisy measurements of the signal are obtained as follows:

$$x(t)=s(t,\boldsymbol{\theta})+n(t), \quad -\frac{T}{2}\leqslant t\leqslant\frac{T}{2}, \tag{39}$$

where $n(t)$ is the noise, assumed Gaussian with a known spectral density denoted by $S_n(f)$.

A. Cramer-Rao bound

The CRB was defined in Eq. (2) to be the inverse of the Fisher information at $\boldsymbol{\theta}$. The latter is a matrix of second derivatives of the likelihood ratio of an observation with respect to $\boldsymbol{\theta}$. In the case of stationary Gaussian noise, the matrix elements reduce to

$$\mathcal{J}_{ij}=\left(\frac{\partial s}{\partial\theta_i}\middle|\frac{\partial s}{\partial\theta_j}\right), \tag{40}$$

where $(s_1|s_2)$ denotes the inner product between two signals, $s_1(t)$ and $s_2(t)$. In terms of the signal's Fourier transforms, $\tilde{s}_1(f)$ and $\tilde{s}_2(f)$, and the spectral density of the noise, $S_n(f)$, the inner product can be expressed as

$$(s_1|s_2)=2\int_0^\infty\frac{\tilde{s}_1^*(f)\,\tilde{s}_2(f)+\tilde{s}_1(f)\,\tilde{s}_2^*(f)}{S_n(f)}\mathrm{d}f. \tag{41}$$

The integral is a measure of the degree of ''overlap'' between the two signals, and in radar applications it is often termed the *ambiguity function*. Thus, Eq. (40) can be interpreted as the local curvature of the signal ambiguity function around its maximum. Numerical integration is generally required to compute the elements of the Fisher information matrix. It is then straightforward to perform the inversion and obtain the CRB. The diagonal elements of the inverted Fisher matrix are the Cramer-Rao bounds on the variances of each of the signal's parameters.

B. Weiss-Weinstein bound

The calculation of the WWB relies upon evaluating the semi-invariant moment generating function $\eta(1/2;\boldsymbol{\delta}_i,\boldsymbol{\delta}_j)$ in Eq. (21). It is not difficult to show (see the Appendix in [28] for details) that, for stationary and Gaussian noise, this function can be reduced to

$$\eta(1/2;\boldsymbol{\delta}_i,\boldsymbol{\delta}_j)=\ln\ C(\boldsymbol{\delta}_i,\boldsymbol{\delta}_j)+\mu(1/2;\boldsymbol{\delta}_i-\boldsymbol{\delta}_j), \tag{42}$$

where

$$C(\boldsymbol{\delta}_i,\boldsymbol{\delta}_j)=\int_\Theta\sqrt{p(\boldsymbol{\theta}+\boldsymbol{\delta}_i)p(\boldsymbol{\theta}+\boldsymbol{\delta}_j)}\mathrm{d}\boldsymbol{\theta}, \tag{43}$$

and the region of integration is $\Theta=\{\boldsymbol{\theta}:p(\boldsymbol{\theta})>0\}$. In Eq. (42), the first term embodies the prior information about the parameters. Consider the single test-point version of the WWB for a single parameter having a uniform prior on the interval $[-D/2,D/2]$. The integral is simple to evaluate and it yields

$$C(\delta)=\ln\left(1-\frac{|\delta|}{D}\right). \tag{44}$$

After some algebra, the second term in Eq. (42) reduces to

$$\mu(1/2;\boldsymbol{\delta}_i-\boldsymbol{\delta}_j)=-\frac{1}{4}\rho^2[1-\gamma(\boldsymbol{\delta}_i-\boldsymbol{\delta}_j)], \tag{45}$$

where $\rho^2\equiv(s|s)$ is the squared amplitude (energy) signal-to-noise ratio, and

$$\gamma(\boldsymbol{\delta}_i-\boldsymbol{\delta}_j)=\frac{(s[\boldsymbol{\theta}+\boldsymbol{\delta}_i-\boldsymbol{\delta}_j]|s[\boldsymbol{\theta}])}{(s[\boldsymbol{\theta}]|s[\boldsymbol{\theta}])} \tag{46}$$

is the normalized signal ambiguity function. It is often the case in practice that the latter function is independent of $\boldsymbol{\theta}$, and then its calculation is greatly simplified. This will be the case for the examples that are presented later. However, numerical integrations are still generally required to compute the signal ambiguity function. Moreover, these integrations have to be performed for every set of test point locations in the parameter space. As the WWB also requires inversion of a matrix having dimension equal to the number of test points, it is clearly desirable to keep the number of test points down to a minimum. An indicator of the number of test points that are required for a given problem, and their optimal locations, is the shape of the signal ambiguity function. As we shall see later, for the coalescing binary waveform this function has a very well-defined shape.

C. Ziv-Zakai bound

The main term on which the evaluation of the ZZB in Eq. (33) hinges is $P_{\min}$, the minimum probability of error in a binary detection problem. An exact expression for $P_{\min}$ exists for the decision problem of discriminating between two *equally likely* signal vectors, $\mathbf{s}_1$ and $\mathbf{s}_2$, in a background of Gaussian noise of covariance $\mathbf{K}$. The minimum probability of error is then given simply by

$$P_{\min}=\Phi\left(\frac{d}{2}\right), \tag{47}$$

where d is the normalized distance between the signals,

$$d=\sqrt{(\mathbf{s}_2-\mathbf{s}_1)^{\mathsf{T}}\mathbf{K}^{-1}(\mathbf{s}_2-\mathbf{s}_1)}, \tag{48}$$

and

$$\Phi(z)=\int_z^\infty\frac{1}{\sqrt{2\pi}}\,e^{-t^2/2}\mathrm{d}t. \tag{49}$$

If the inner products under the square root sign in Eq. (48) are evaluated, one finds that

$$d=\sqrt{-\mu(1/2;\boldsymbol{\delta}_i-\boldsymbol{\delta}_j)}, \tag{50}$$

where μ is given by Eq. (45). Therefore the calculation of the ZZB, like the WWB, is crucially dependent on the shape of the signal ambiguity function. In the case of the WWB this dictates the number of test points and their locations in order to achieve a tight bound. In terms of the ZZB, the shape of the signal ambiguity function defines a path of integration in Eq. (33), subject to the constraint under which the integration is evaluated.

VI. ILLUSTRATIVE EXAMPLE

In this section we apply the CRB and one of the Bayesian bounds (WWB) to a simple scalar parameter estimation problem. Our intention is to investigate the conditions under which Bayesian bounds are tighter than local bounds.

A common feature of all the bounds that we have presented is that they depend on the shape of the signal ambiguity function, γ, rather than the signal shape, when the background of noise is Gaussian. Often the shape of the signal ambiguity function is the same for all underlying values of the signal's parameters, greatly simplifying the calculation of the bounds.

The CRB only probes the shape of the signal ambiguity function around its maximum. Structure in the ambiguity function away from the maximum could be enhanced by noise and masquerade as a false peak. This would confound a maximum likelihood parameter estimator, and may lead to numerical parameter estimation errors that depart significantly from the theoretical CRB.

While the CRB is ''blind'' to the presence of sidebands in the signal ambiguity function, the WWB and ZZB are able to capture this structure. In the case of the WWB this is achieved through the test points. As well as probing around the main lobe of the ambiguity function, test points may also be placed around the secondary maxima. Similarly, the ZZB is generally tighter than the CRB if a path of integration is selected to traverse all of the predominant lobes in γ.

The difference between the CRB and the WWB is best illustrated through an example. We consider two signals, s_1 and s_2, characterized by a scalar parameter, θ, that enters the signals nonlinearly. The signal waveforms need not concern us here, only their ambiguity functions $\gamma_i(\theta,\Delta\theta)$, where $i=1,2$. We have chosen signals whose γ's are independent of θ and depend only on displacements in θ, i.e. $\Delta\theta$. These ambiguity functions are displayed in Fig. 2. The signal s_1 was designed so that γ_1 has only a single broad maximum. The other signal, s_2, has γ_2 comprised of a number of significant secondary lobes. Note also that γ_1 is actually the ''envelope'' of γ_2.

We assume that θ has a uniform prior. However, the region of support of the prior is set much larger than the anticipated parameter estimation errors, so that the prior does not actually impact upon the estimation accuracy for this problem.

For s_1, the WWB was computed by placing N test points uniformly along the lobe of the signal ambiguity function. It was found that the resulting bound was not very sensitive to where the test points were placed along the lobe for this signal. A variable number of test points (up to 20) were used to study the convergence of the bound. This was attained for only 4 test points. For s_2, the test points were placed around the main lobe of the ambiguity function and also around the principal secondary maxima. In fact only the first three secondary lobes needed to be covered: Again the WWB exhibited rapid convergence.

The CRB was also obtained for s_1 and s_2 over an identical range of SNR. This was obtained by inverting the Fisher matrix (40) for this specific problem. However, the CRB can also be computed in terms of the WWB formalism for a single test point, δ, allowing $\delta\to 0$. It should be remarked

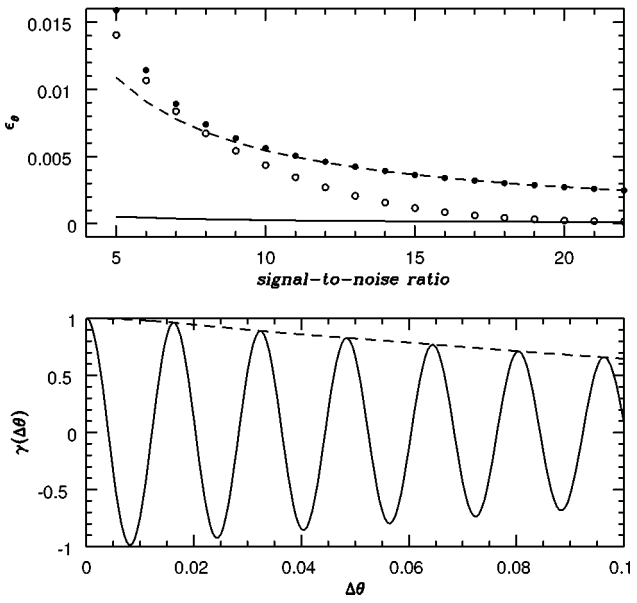

FIG. 2. This diagram illustrates the link between the structure of the ambiguity function and the accuracy of parameter estimation. The lower panel displays the ambiguity function for two different one-parameter signals, as a function of the displacement $\Delta\theta$ from the true parameter value: s_1 (dashed line) was chosen to have an ambiguity function with one broad maximum; s_2 (bold line) was designed to have a more structured ambiguity function with many secondary sidelobes. For the special signals considered here, both ambiguity functions depend only on $\Delta\theta$, and not on the actual value θ. In the upper panel we display the results of applying the Cramer-Rao (dashed and bold line for s_1 and s_2, respectively) and the Weiss-Weinstein theory (solid and open circles for s_1 and s_2, respectively) to bound the mean-square error ϵ_θ in the estimation of the signal's parameter as a function of the signal-to-noise ratio. See text for further details.

that in the multiple test-point formulation of the WWB, one test point is always forced to have $\delta\to 0$. This ensures that the WWB reduces to the CRB at a large SNR.

The bounds, $\epsilon_\theta^{(i)}$ $(i=1,2)$, that we calculated are displayed in the top panel of Fig. 2. For s_1, the WWB and the CRB are in good agreement down to SNR $\simeq 11$. At smaller values of SNR, the WWB is a tighter bound than the CRB but not significantly so. This result is not too surprising: γ_1 has only a single broad maximum and the CRB probes the curvature of this lobe. Similarly the WWB probes the structure in the ambiguity lobe, although it is able to probe further away from the maximum with respect to the CRB. As the lobe is broad, the WWB is generally a little tighter than the CRB. The results for s_2 are significantly different. Here, the WWB departs from the CRB at a much higher value of the SNR, around 20. This is because the CRB is blind to the secondary maxima in γ_2 that the WWB is able to capture through a judicious choice of test points. The discrepancy between the CRB and the WWB for this example is striking: a factor of ~ 5 at SNR$=10$ and more than an order of magnitude at SNR$=5$. The WWB falls significantly as the SNR increases, while the CRB remains fairly constant, due to the sharp maximum in γ_2.

It is also interesting to compare the behavior of $\epsilon_\theta^{(1)}$ and $\epsilon_\theta^{(2)}$. At a high SNR the accuracy in the determination of θ

for s_2 is better than for s_1. This is clearly due to the sharp maximum in γ_2. At a low SNR, we expect $\epsilon_\theta^{(1)} \sim \epsilon_\theta^{(2)}$ because the ambiguity functions have roughly the same ''global'' profile. This intuition is borne out in the results of the Bayesian analysis (WWB), but not for the local analysis (CRB) where $\epsilon_\theta^{(2)} < \epsilon_\theta^{(1)}$ at all SNR's.

VII. APPLICATION TO COALESCING BINARY PARAMETER ESTIMATION

We are now in a position to investigate whether Bayesian bounds provide a tighter constraint than the CRB on the error covariance matrix for the parameters of a gravitational wave signal generated during the inspiral phase of a compact object binary. In the following we use units where $G=c=1$. This implies a conversion factor $1\mathrm{M}_\odot = 4.926\times 10^{-6}$ s.

A. Signal and noise model

The coalescing binary inspiral waveform can be cast in the following generic form:

$$h(t) = A[\pi f(t)]^{2/3}\cos[\phi(t)], \tag{51}$$

where $f(t)$ is the instantaneous gravitational wave frequency, $\phi(t)$ the instantaneous phase, and A the amplitude. We will consider the phasing of the wave, as well as the amplitude evolution, only up to Newtonian order. In this approximation the factor A is a constant, complicated, function of the binary's distance, location in the sky, *chirp mass* $\mathcal{M} \equiv m_1^{3/5} m_2^{3/5}/(m_1+m_2)^{1/5}$ (where m_1 and m_2 are the masses of the compact objects) and the detector's antenna pattern [5]. Its precise functional form need not concern us further. Frequency and phase read

$$f(t) = f_a\left[1 - \frac{t-t_a}{\tau}\right]^{8/3} \tag{52}$$

and

$$\phi(t) = \frac{16\pi f_a \tau}{5}\left\{1 - \left[\frac{f(t)}{f_a}\right]^{-5/3}\right\} + \Phi_a, \tag{53}$$

where the constant τ, sometimes referred to as the chirp time, can be cast in terms of the chirp mass of the binary as

$$\tau = \frac{5}{256}\mathcal{M}^{-5/3}(\pi f_a)^{-8/3}. \tag{54}$$

The constants f_a and Φ_a are respectively the frequency and phase of the signal at the arbitrary time $t=t_a$. The waveform is characterized in terms of 3 parameters t_a, Φ_a, and τ. The amplitude parameter A enters the waveform linearly and it will be incorporated later into our definition of signal-to-noise ratio [see Eq. (62)]. Of course, the parametrization of the signal is not unique and one can express $h(t)$ as a function, for example, of the time to coalescence and the phase of the wave at this time. In fact, the latter parametrization may have some advantages. However, our main goal here is to assess the *relative* difference between the CRB and Bayesian bounds and so this detail of the parametrization is not crucial: We require only consistency in the choice of parametrization of the signal in order to compare the bounds. In particular the set of parameters that we are assuming here may not correspond to physical ones, and this would be the case if we extended this setting to post-Newtonian waveforms [27]. However, there is one crucial feature of the wave parametrization that we adopt here: It produces a signal ambiguity function that is independent of the intrinsic values of the parameters and depends only upon their displacements. This fact is more transparent if we examine the signal's Fourier transform $\tilde{h}(f)$, which in the stationary phase approximation [18] reads

$$\tilde{h}(f) = \mathcal{N} f^{-7/6}\exp[\Psi(f)], \tag{55}$$

where

$$\mathcal{N} = A\pi^{2/3}\left(\frac{2\tau}{3}\right)^{1/2} f_a^{4/3} \tag{56}$$

is a normalization constant, and

$$\Psi(f) = i\sum_{\nu=1}^{3}\psi_\nu(f)\lambda^\nu - i\frac{\pi}{4}; \tag{57}$$

λ^ν represents the parameter vector

$$\lambda^\nu \equiv (t_a, \Phi_a, \tau), \tag{58}$$

and

$$\begin{aligned}\psi_1 &= 2\pi f,\\ \psi_2 &= -1,\\ \psi_3 &= 2\pi f - \frac{16\pi f_a}{5} + \frac{6\pi f_a}{5}\left(\frac{f}{f_a}\right)^{-5/3}.\end{aligned} \tag{59}$$

Notice that the signal's parameters enter linearly into the phase (57) of its Fourier transform. The normalized ambiguity function for the signal (55) is given by

$$\gamma(\Delta\lambda^\nu) = J^{-1}\int \frac{f^{-7/3}}{S_n(f)}\cos\left\{\mathrm{Re}\left[\sum_{\nu=1}^{3}\psi_\nu(f)\Delta\lambda^\nu\right]\right\}\,\mathrm{d}f, \tag{60}$$

where

$$J = \int \frac{f^{-7/3}}{S_n(f)}\,\mathrm{d}f \tag{61}$$

and the integral is defined over the frequency interval, within the instrument's sensitivity band, spanned by the signal. The optimal signal-to-noise ratio reads

$$\rho^2 \equiv (h|h) = 4\mathcal{N}^2 J \tag{62}$$

and, therefore, incorporates the amplitude parameter A via the definition of $\mathcal{N}$; cf. Eq. (56).

Our model for the noisy spectral density, $S_n(f)$, is intended to be representative of the performance of the first

stage LIGO detectors. An analytic fitting formula for this has been presented in [31], and we utilize it here. Accordingly to [27], we have considered the observational window confined to the frequency interval 40–750 Hz. We suppose that the final frequency of inspiral is outside the considered bandwidth, so that the integral involved in the definition of the inner product (41) is evaluated on the same frequency range.

B. Calculating the bounds

In order to compare the bounds on parameter estimation errors given by local and global approaches, we computed the CRB, WWB and ZZB. The computational steps are described here, focusing particularly on the WWB and the ZZB. Our discussion is centered upon the evaluation of bounds for the time-of-arrival parameter, t_a $(=\lambda_1)$. Similar results apply to Φ_a and τ.

The CRB involves the computation of the diagonal elements of the inverse of the Fisher information matrix (40), that is,

$$\epsilon_i^2 = \mathcal{J}_{ii}^{-1}; \tag{63}$$

for the signal (55) and the parameters (58), we have

$$\frac{\partial \tilde{h}}{\partial \lambda^j} = i\mathcal{N} f^{-7/6} \psi_j(f) \exp[\Psi(f)] \tag{64}$$

and, therefore,

$$\mathcal{J}_{jk} = \frac{\rho^2}{J} \int \psi_j(f)\psi_k(f) \frac{f^{-7/3}}{S_n(f)}\, df. \tag{65}$$

The evaluation of Eq. (63) is straightforward from Eqs. (59) and (65) and has been thoroughly studied in many papers [18,19,32–34], where further details can be found.

The WWB involves the computation of Eq. (19) and therefore of the $M\times N$ matrix $\mathcal{H}$ and the $N\times N$ matrix $\mathcal{G}$, where M is the number of parameters (3 in this problem) and N is the number of test points $\{\boldsymbol{\delta}_j, j=1,\ldots,N\}$. The test points are now given explicitly by $\Delta\lambda_j^\nu$ (the lower and upper indices labelling the test points and the parameters, respectively). The elements of the matrix $\mathcal{H}$ are therefore $\mathcal{H}_{\nu j} = \Delta\lambda_j^\nu$.

We will assume here that λ^ν has a uniform prior distribution with a region of support that is much larger than the anticipated errors on the parameters. Therefore the prior will not impact on the calculation of the WWB.

Equations (20) and (45) read now

$$\mathcal{G}_{jk} = 2\frac{\exp[\rho^2\gamma(\Delta\lambda_j^\nu - \Delta\lambda_k^\nu)/4] - \exp[\rho^2\gamma(\Delta\lambda_j^\nu + \Delta\lambda_k^\nu)/4]}{\exp[\rho^2\gamma(\Delta\lambda_j^\nu)/4]\exp[\rho^2\gamma(\Delta\lambda_j^\nu)/4]} \tag{66}$$

and

$$\eta(1/2;\Delta\lambda_j^\nu, \Delta\lambda_k^\nu) = \mu(1/2;\Delta\lambda_j^\nu - \Delta\lambda_k^\nu). \tag{67}$$

A crucial issue for a reliable computation of the bound is the placing of test points: (i) In order to get the CRB in the limit of a high SNR, the elements $\mathcal{H}_{\nu j}$ for $j=1,2,3$ (notice that j runs from 1 to N) have been chosen according to $\mathcal{H}_{\nu j} = \delta_{\nu\alpha}\Delta\lambda_j^\alpha$, with $\Delta\lambda_j^\alpha \ll 1$ (here $\delta_{\nu\alpha}$ is the Kronecker symbol); indeed, we always probe the primary peak of the ambiguity function, as does the CRB; (ii) in order to get the tightest possible bound at a low SNR, we placed the other test points ($\mathcal{H}_{\nu j}$ for $j>3$) along the maxima of the ambiguity function, using the test-case problem presented in Sec. VI as a guideline. In Fig. 3 we show $\gamma(\Delta\lambda^\nu)$, in the plane $(\Delta t_a, \Delta\tau)$, maximized with respect to $\Delta\Phi_a$ (we already know that the regions where the ambiguity function is small do not contribute significantly to the result). The plot is enlightening as γ consists of a long, sharp ridge and clearly indicates that the test points need to be spread along that curve. We placed up to 25 points (about the maximum permitted by the numerical routines implemented for the matrix inversion) with different choices of their separation and distance from the origin. We noted, in fact, that with only 4 almost equally spaced points (spacing $\simeq 16$ ms), the result did not change significantly, in agreement with what was found in the toy problem.

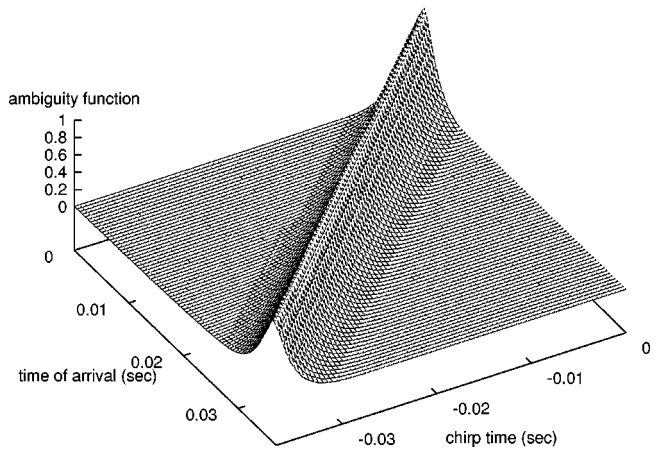

FIG. 3. The ambiguity function, maximized with respect to the phase of arrival Φ_a, for the signal (55) as a function of the time of arrival and of the chirp time.

The evaluation of the ZZB involves the computation of the integral (33), using the minimum probability error given by Eqs. (47) and (50). As we have stressed before, the strategy of computation replaces here the spreading of test points with the selection of the integration path. Our discussion of the evaluation of the WWB indicates that the integration has to be performed along the ridge of the ambiguity function shown in Fig. 3, in order to produce the tightest bound. Of course, no ''valley filling'' function was needed, as γ is a smooth curve free from oscillations for the signal that we were studying. We carried out the integration up to a maximum displacement from the origin $\simeq 0.2$ s (of the same order of the position of the last test point during the investigation of the WWB), after which no appreciable improvement was found. As in the case of the computation of the WWB, the prior probability on λ_1 did not impact on the bound, because the region of support was chosen to be much larger than the anticipated error.

C. Results

The bounds calculated following the three different theoretical approaches (CRB, WWB and ZZB) were computed

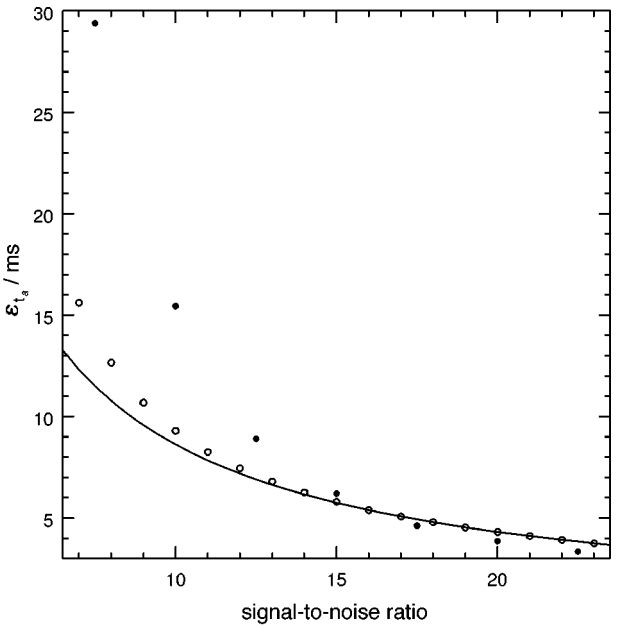

FIG. 4. Comparison between local and global theoretical bounds on the time-of-arrival error with actual parameter estimation errors obtained by applying the maximum likelihood method to simulated data (bold line, CRB; open circles, maximum of WWB and ZZB; solid circles, maximum likelihood error).

for values of ρ in the relevant range $7 \leqslant \rho \leqslant 25$. In the same SNR interval we compared these results with those obtained by means of numerical simulations [27] that implement maximum likelihood estimators. The root-mean-square error bounds on the time-of-arrival parameter t_a are displayed as a function of ρ in Fig. 4. All the bounds converge to the CRB at $\rho \sim 15$, and at this SNR the CRB is attained by the maximum likelihood estimator. At smaller values of SNR, the Bayesian bounds deviate from the CRB, providing a slightly tighter result ($\simeq 6\%$ at $\rho = 10$ and $\simeq 25\%$ at $\rho = 7$). This not too severe discrepancy can be explained by the structure of γ: Both the local and global approaches probe the ambiguity function around its origin, but the Bayesian bound is able to follow it further away from the origin (see the discussion in Sec. VI). The behaviors of WWB and ZZB were found to be very similar, although not exactly equal: The SNR threshold at which they depart from the CRB is $\rho \simeq 14$, for the ZZB, and $\rho \simeq 12$, for the WWB, but the latter provides a better constraint at low SNR's. This agrees with results from other applications of the bound to time-of-arrival parameter estimation problems in radar and in sonar applications (see [30] and references therein). The striking feature of the comparison is given by the maximum likelihood errors obtained in numerical experiments: While matching the behavior of (local and global) lower bounds at a high SNR ($\rho \gtrsim 15$), they produce errors that are dramatically higher than expected theoretically at low SNR's: about 65% at $\rho = 10$ and more than a factor of 2 at $\rho = 7$.

VIII. CONCLUSIONS

We have reassessed the issue of information extraction with respect to observations of coalescing binaries by interferometric gravitational wave detectors. After discussing the main properties of global and local bounds on parameter estimation, we applied a set of these bounds to a gravitational wave parameter estimation problem. In particular we have introduced the Weiss-Weinstein and the Ziv-Zakai bounds. These provide fundamental lower limits on the mean-square error on the parameters that describe a signal, independently of the actual estimator that is adopted in the data-analysis process. In addition these bounds easily incorporate any *a priori* information that is available about the problem and do not suffer from limitations that affect local bounds and the Bayesian version of the Cramer-Rao bound. In short, these global bounds can be used to benchmark the performance of any practical information extraction technique.

We have applied the bounds to the case of laser interferometric measurements of waveforms that are characteristic of those emitted by inspiraling compact binaries in a background of noise that is characteristic of the performance of first-stage detectors. Comparisons between the Cramer-Rao, Weiss-Weinstein and Ziv-Zakai bounds on the MSE and actual maximum likelihood errors obtained by numerical experiments, over a wide range of SNR, show that (i) at high signal-to-noise ratios (SNR $\gtrsim 15$) all the approaches converge to the same value of the MSE. In this regime one can regard the Fisher information matrix as a simple and reliable tool to compute ''realistic'' bounds on estimation errors. Maximum likelihood methods are probably adequate to extract astrophysical information from noisy data at these SNR's (although a definitive statement is premature, pending a detailed analysis applied to more general waveforms). (ii) At low signal-to-noise ratios (SNR $\lesssim 10$, in which most of the events are likely to be recorded during the first years of operation of the detectors) the WWB and ZZB produce a more stringent constraint ($\simeq 25\%$ at $\rho = 7$) on the MSE with respect to the CRB, indicating that the latter can underestimate the errors in this regime. Perhaps more seriously, all the bounds are about 2 times smaller than the errors that are obtained in the numerical experiments.

This analysis suggests that maximum likelihood techniques need to be refined, or complemented, in order to attain the lowest possible value of the errors. Our study of toy problems and the results from previous investigations in different fields suggest that these are within a few percent of those predicted theoretically via the WWB and ZZB. A first attempt toward the understanding of the outcome of numerical experiments and the performances of maximum likelihood estimators has been recently reported in [40]. We are currently exploring the possibility of implementing conditional-mean estimators to improve parameter estimation accuracy, and will report on this in a forthcoming paper [41]. The conditional mean estimator is generally intractable to implement as it requires the computation of a multidimensional integral over the space of all the parameters (in realistic cases more than ten), even though suitable strategies to reduce the amount of computation have been proposed and successfully tested in (simple) cases [42]. Hierarchical strategies combining maximum likelihood and non-linear filtering could also speed up the process.

Finally, it is important to underscore the point that the form in which we have presented the WWB and ZZB is completely general. It can be applied to parameter estimation

problems for other kinds of signals (e.g. pulsars) and other instruments and/or arrays of detectors.

ACKNOWLEDGMENTS

We are indebted to Kristine Bell for sending us her thesis dissertation on the ZZB and graciously answering many of our subsequent queries. In addition, we thank Tony Weiss for fielding questions and providing references to the WWB. We acknowledge helpful discussions with Kostas Kokkotas, Andrzej Krolak, Bernard Schutz, and B. Sathyaprakash. A.V. acknowledges financial support from the Fondazione Della Riccia and the Department of Physics and Astronomy of the UWCC for the kind hospitality. The last months of A.V.'s contribution to this work were supported by the Max Planck Gesellschaft.

[1] A. Abramovici *et al.*, Science **256**, 325 (1992).
[2] C. Bradaschia *et al.*, Nucl. Instrum. Methods Phys. Res. A **289**, 518 (1990).
[3] K. Danzmann, in *Gravitational Wave Experiments*, edited by E. Coccia, G. Pizzella and F. Ronga (World Scientific, Singapore, 1995), pp. 100–111.
[4] K. Tsubono, in *Gravitational Wave Experiments*, edited by E. Coccia, G. Pizzella and F. Ronga (World Scientific, Singapore, 1995), pp. 112–114.
[5] K. S. Thorne, in *300 Years of Gravitation*, edited by S. Hawking and W. Israel (Cambridge University Press, Cambridge, England, 1987), pp. 330–458.
[6] C. Cutler *et al.*, Phys. Rev. Lett. **70**, 2984 (1993).
[7] K. S. Thorne, in *Proceedings of the Snowmass 95 Summer Study on Particle and Nuclear Astrophysics and Cosmology*, edited by E.W. Kolb and R. Peccei (World Scientific, Singapore, 1995).
[8] C. M. Will, in *Proceedings of the 8th Nishinomiya-Yukawa Symposium on Relativistic Cosmology*, edited by M. Sasaki (Universal Academic Press, Japan, 1994).
[9] Numerical Relativity Grand Challenge Alliance, 1995. References and information on the WWW at http://jean-luc.ncsa.uiuc.edu/GC.
[10] L. Blanchet, ''Gravitational radiation from relativistic sources,'' presented at the Les Houches School on Gravitational Radiation, gr-qc/9607025.
[11] R. Narayan, T. Piran, and A. Shemi, Astrophys. J. Lett. **379**, L17 (1991).
[12] E. S. Phinney, Astrophys. J. **380**, L17 (1991).
[13] E. P. J. Van den Heuvel and D. R. Lorimer, Mon. Not. R. Astron. Soc. **283**, L37 (1996).
[14] V. M. Lipunov, K. A. Postnov, and M. E. Prokhorov, New Astron. (to be published).
[15] V. M. Lipunov, K. A. Postnov, and M. E. Prokhorov, Astron. Lett. (to be published).
[16] C. M. Will, Phys. Rev. D **50**, 6058 (1994).
[17] E. Poisson, Phys. Rev. D **54**, 5939 (1996).
[18] C. Cutler and É. Flanagan, Phys. Rev. D **49**, 2658 (1994).
[19] E. Poisson and C. M. Will, Phys. Rev. D **52**, 848 (1995).
[20] L. S. Finn, Phys. Rev. D **53**, 2878 (1996).
[21] B. F. Schutz, Nature (London) **323**, 310 (1986).
[22] D. Marković, Phys. Rev. D **48**, 4738 (1993).
[23] D. F. Chernoff and L. S. Finn, Astrophys. J. **411**, L5 (1993).
[24] Y. Wang and E. L. Turner, Phys. Rev. D **56**, 724 (1997).
[25] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, NJ, 1946).
[26] H. L. Van Trees, *Detection Estimation and Modulation Theory* (Wiley, New York, 1968), Pt. I.
[27] R. Balasubramanian, B. S. Sathyaprakash, and S. V. Dhurandhar, Phys. Rev. D **53**, 3033 (1996).
[28] E. Weinstein and A. J. Weiss, IEEE Trans. Inf. Theory **34**, 338 (1988).
[29] J. Ziv and M. Zakai, IEEE Trans. Inf. Theory **15**, 386 (1969).
[30] K. L. Bell, Ph.D. thesis, George Mason University, 1995.
[31] L. S. Finn and D. F. Chernoff, Phys. Rev. D **47**, 2198 (1993).
[32] P. Jaranowski and A. Krolak, Phys. Rev. D **49**, 1723 (1994).
[33] A. Krolak, K. D. Kokkotas, and G. Schäfer, Phys. Rev. D **52**, 2089 (1995).
[34] P. Jaranowski, A. Krolak, K. D. Kokkotas, and G. Tsegas, Class. Quantum Grav. **13**, 1279 (1996).
[35] C. W. Helström, *Statistical Theory of Signal Detection*, 2nd ed. (Pergamon, London, 1968).
[36] E. W. Barankin, Ann. Math. Stat. **20**, 477 (1949).
[37] M. H. D. Davis, in *Gravitational Wave Data Analysis*, edited by B.F. Schutz (Kluwer, Dordrecht, 1989), pp. 73–94.
[38] S. Bellini and G. Tartara, IEEE Trans. Commun. **22**, 340 (1974).
[39] B. F. Schutz, in *Data Processing Analysis and Storage for Interferometric Antennas*, edited by D. G. Blair (Cambridge University Press, Cambridge, United Kingdom, 1991), pp. 406–452.
[40] R. Balasubramanian and S. V. Dhurandhar, ''Gravitational waves from coalescing binaries: Estimation of parameters'' gr-qc/9702015; ''Estimation of parameters of gravitational waves from coalesing binaries,'' gr-qc/9708003.
[41] A. Vecchio and D. Nicholson (in progress).
[42] A. Pasetti, M.Sc. thesis, University of London, 1987.