Rapid Bayesian position reconstruction for gravitational-wave transients

Leo P. Singer and Larry R. Price

Phys. Rev. D **93**, 024013 — Published 14 January 2016

DOI: 10.1103/PhysRevD.93.024013

WHOOMP! (There It Is)
Rapid Bayesian position reconstruction for gravitational-wave transients

Leo P. Singer* and Larry R. Price†
LIGO Laboratory, California Institute of Technology, Pasadena, CA 91125, USA

Within the next few years, Advanced LIGO and Virgo should detect gravitational waves (GWs) from binary neutron star and neutron star–black hole mergers. These sources are also predicted to power a broad array of electromagnetic transients. Because the electromagnetic signatures can be faint and fade rapidly, observing them hinges on rapidly inferring the sky location from the gravitational wave observations. Markov chain Monte Carlo (MCMC) methods for gravitational-wave parameter estimation can take hours or more. We introduce BAYESTAR, a rapid, Bayesian, non-MCMC sky localization algorithm that takes just seconds to produce probability sky maps that are comparable in accuracy to the full analysis. Prompt localizations from BAYESTAR will make it possible to search electromagnetic counterparts of compact binary mergers.

PACS numbers: 04.80.Nn, 04.30.Tv, 02.50.Tt

The Laser Interferometer GW Observatory (LIGO; [1, 2]) has just begun taking data [3] in its 'Advanced' configuration. The two LIGO detectors will ultimately increase their reach in volume within the local Universe by three orders of magnitude as compared to their initial configurations through 2010. They form the first parts of a sensitive global GW detector network, soon to be augmented by Advanced Virgo [4] and later by the Japanese KAGRA facility [5, 6] and LIGO–India [7].

The most readily detectable sources of GWs include binary neutron star mergers, with 0.4–400 events per year within the reach of Advanced LIGO at final design sensitivity [8]. These binary systems are not only efficient sources of GWs, but also potential sources of detectable electromagnetic (EM) transients from the aftermath of the tidal disruption of the neutron stars (NSs). Metzger and Berger [9] argue that the most promising EM counterparts are the hypothesized optical/infrared "kilonovae" powered by the radioactivate decay of r-process elements synthesized within the neutron-rich ejecta. These are expected to be faint, red, and peak rapidly, reaching an absolute magnitude of only $M_R \sim -13$ within a week, though rising several magnitudes brighter in the infrared [10].

Several mechanisms could make the kilonovae brighter, bluer, and hence more readily detectable [11, 12], but peak even earlier, within hours. If, as is widely believed [13–16], binary neutron star (BNS) mergers are indeed progenitors of short gamma-ray bursts (GRBs), then a small (due to jet collimation) fraction of Advanced LIGO events could also be accompanied by a bright optical afterglow, but this signature, likewise, would peak within hours or faster.

Adding to the challenge of detecting a faint, short-lived optical transient, there is an extreme mismatch between the sky localization accuracy of GW detector networks, $\sim$10–500 deg^2 [17–27], and the fields of view (FOVs) of 1–8 m-class optical telescopes. Wide-field optical transient facilites such as BlackGEM (0.6 m/2.7 deg^2), the Zwicky Transient Facility (ZTF; 1.2 m/47 deg^2) [28], the Dark Energy Camera (DECam; 4 m/3 deg^2), or the Large Synoptic Survey Telescope (LSST; 8.4 m/9.6 deg^2), operated in "target of opportunity" mode, may be able to tile these large areas rapidly enough to find the one needle in the haystack that is connected to the GW event. However, prompt and accurate GW position reconstructions will be of the utmost importance for guiding the selection of fields to observe.

The final science run of the initial LIGO and Virgo instruments saw the first joint search for GW and electromagnetic emission from compact binaries. This involved several advances in the GW data analysis [29], including the first real-time matched-filter detection pipeline (MBTA, Multi-Band Template Analysis; [30]), a semi-coherent, ad hoc rapid triangulation code (Timing++), and the first version of a rigorous Bayesian MCMC parameter estimation code (LALINFERENCE; Aasi *et al.* 31)—all in service of the first search for X-ray [32] and optical [33] counterparts of GW triggers, by a consortium of facilities. Despite the technical achievements in the GW data analysis, there was an undesirable tradeoff between the speed as well as accuracy of the rapid localization and the full parameter estimation: the former could analyze a detection candidate in minutes, whereas the latter took half a day; the latter decreased the area on the sky by a factor of 1/20 over the former but took 1000 times as long to run [26].

The success of EM follow-up of LIGO events will depend critically on disseminating high quality sky localizations within a timescale of minutes to hours. To that end, we have devised a rapid and accurate Bayesian sky localization method that takes mere seconds to achieve approximately the same accuracy as the full MCMC analysis. Our key insights are the following:

1. Nearly all of the information in the GW time series that is informative for sky localization is encap-

* leo.singer@ligo.org
† larryp@caltech.edu

sulated within the matched-filter estimates of the times, amplitudes, and phases on arrival at the detectors. To infer the position and distance of a GW event, we only have to consider three numbers per detector rather than a densely sampled strain time series per detector.

2. The matched-filter pipeline can be treated as a measurement system in and of itself. Just like the strain time series from the detectors, the resultant times, amplitudes, and phases have a predictable and quantifiable measurement uncertainty that can be translated into a likelihood function suitable for Bayesian inference.

3. The Fisher information matrix will provide clues as to suitable forms of this likelihood function. Recent GW parameter estimation literature has largely rejected the Fisher matrix,[1] but this is mostly on the grounds of the abuse of the related Cramér–Rao lower bound (CRLB) outside its realm of validity (i.e., at low to moderate signal-to-noise ratio (S/N)) [35–38]. However, we recognize that the block structure of the Fisher matrix provides important insights, and is a useful quantity for checking the validity of the aforementioned likelihood function, quite independent of the CRLB.

4. The Fisher matrix teaches us that errors in sky localization are semi-independent from errors in masses. This implies that if we care only about position reconstruction and not about jointly estimating masses as well, then we can reduce the dimensionality of the parameter estimation problem significantly. Moreover, this frees us of the need to directly compute the expensive post-Newtonian model waveforms, making the likelihood itself much faster to evaluate.

5. Thanks to a simple likelihood function and a well-characterized parameter space, we may dispense with costly and parallelization-resistant MCMC integration, and instead perform the Bayesian marginalization with classic, deterministic, very low order Gaussian quadrature.

6. The Bayesian inference scheme thus designed to operate on the matched-filter detection pipeline outputs could be trivially generalized to operate on the full GW time series within the same computational constraints. This would yield a fast and coherent localization algorithm that is mathematically equivalent to the full MCMC parameter estimation, restricted to extrinsic parameters (sky location, binary orientation, and distance).

[1] Though not entirely; see [34].

We call this algorithm BAYESian TriAngulation and Rapid localization (BAYESTAR)[2,3]. It is as fast as Timing++, but nearly as accurate as the rigorous full parameter estimation. It is unique in that it bridges the detection and parameter estimation of GW signals, two tasks that have until now involved very different numerical methods and time scales. Beginning with the first Advanced LIGO observing run, BAYESTAR is providing localizations within minutes of the detection of any BNS merger candidate, playing a key role in enabling rapid follow-up observations.

This paper is organized as follows. In Section I, we describe the GW signal model and sketch the standard detection algorithm, the matched filter bank. In Section II, we describe Bayesian inference formalism and the prevailing method for inferring the parameters of detected candidates, MCMC sampling. In Section III, we propose the BAYESTAR likelihood as a model for the uncertainty in the matched-filter parameter estimates, and discuss its relationship to and consistency with the likelihood for the full GW data. In Section IV, we describe the input to BAYESTAR supplied by the detection pipeline, and the prior distribution on parameters. In Section V, we explain the integration scheme by which the posterior probability distribution is calculated for a given sky location. In Section VI, we show a scheme whereby the sky posterior is sampled on an adaptive HEALPix grid. In Section VIII E, we report the running time of the algorithm on the hardware available on the LIGO Data Grid. Finally, in Section VIII, we quantify the sky localization performance on a comprehensive set of simulated GW events.

I. SIGNAL MODEL AND DETECTION

In the time domain (TD), the strain observed by a single GW interferometer is

$$y_i(t) = x_i(t;\boldsymbol{\theta}) + n_i(t). \tag{1}$$

In the frequency domain (FD),

$$Y_i(\omega) = \int_{-\infty}^{\infty} y(t) e^{-i\omega t} dt = X_i(\omega;\boldsymbol{\theta}) + N_i(\omega), \tag{2}$$

where $X_i(\omega;\boldsymbol{\theta})$ is the GW signal given a parameter vector $\boldsymbol{\theta}$ that describes the GW source, and $N_i(\omega)$ is that detector's Gaussian noise with one-sided power spectral density (PSD) $S_i(\omega) = E\left[|N_i(\omega)|^2\right] + E\left[|N_i(-\omega)|^2\right] =$

[2] A pun on the Cylon battleships in the American television series Battlestar Galactica. The defining characteristic of the Cylons is that they repeatedly defeat humanity by using their superhuman information-gathering ability to coordinate overwhelming forces.

[3] We do not like to mention the final 'L' in the acronym, because then it would be pronounced BAYESTARL, which sounds stupid.

$2E\left[|N_i(\omega)|^2\right]$. We shall denote the combined observation from a network of detectors as $\mathbf{Y}(\omega) \equiv \{Y_i(\omega)\}_i$.

Under the assumptions that the detector noise is Gaussian and that the noise from different detectors are uncorrelated, the likelihood of the observation, $\mathbf{y}$, conditioned upon the parameters $\boldsymbol{\theta}$, is a product of Gaussian distributions:

$$\mathcal{L}(\mathbf{Y};\boldsymbol{\theta}) = \prod_i p(Y_i|\boldsymbol{\theta}) \propto \exp\left[-\frac{1}{2}\sum_i \int_0^\infty \frac{|Y_i(\omega) - X_i(\omega;\boldsymbol{\theta})|^2}{S_i(\omega)}\, d\omega\right]. \quad (3)$$

A compact binary coalescence (CBC) source is specified by a vector of extrinsic parameters describing its position and orientation, and intrinsic parameters describing the physical properties of the binary components[4]:

$$\boldsymbol{\theta} = \begin{bmatrix} \alpha & \text{right ascension} \\ \delta & \text{declination} \\ r & \text{distance} \\ t_\oplus & \text{arrival time at geocenter} \\ \iota & \text{inclination angle} \\ \psi & \text{polarization angle} \\ \phi_c & \text{coalescence phase} \\ \hline m_1 & \text{first component's mass} \\ m_2 & \text{second component's mass} \\ \mathbf{S}_1 & \text{first component's spin} \\ \mathbf{S}_2 & \text{second component's spin} \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}a\\a\\a\\a\\a\\a\\a\end{matrix}}\right\} \text{extrinsic parameters, } \boldsymbol{\theta}_{\text{ex}} \\ \left.\vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}}\right\} \text{intrinsic parameters, } \boldsymbol{\theta}_{\text{in}}. \end{matrix} \quad (4)$$

Assuming a nonprecessing circular orbit, we can write the GW signal received by any detector as a linear combination of two basis waveforms, H_0 and $H_{\pi/2}$ [45]. H_0 and $H_{\pi/2}$ are approximately "in quadrature" in the same sense as the cosine and sine functions, being orthogonal and out of phase by $\pi/2$ at all frequencies. If H_0 and $H_{\pi/2}$ are Fourier transforms of real functions, then $H_0(\omega) = H_0^*(-\omega)$ and $H_{\pi/2}(\omega) = H_{\pi/2}^*(-\omega)$, and we can write

$$H_{\pi/2}(\omega) = H_0(\omega) \cdot \begin{cases} -i & \text{if } \omega \geq 0 \\ i & \text{if } \omega < 0 \end{cases}. \quad (5)$$

For brevity, we define $H \equiv H_0$ and write all subsequent equations in terms of the H basis vector alone. Then, we can write the signal model in a way that isolates all dependence on the extrinsic parameters, $\boldsymbol{\theta}_{\text{ex}}$, into a few coefficients, and all dependence on the intrinsic parameters, $\boldsymbol{\theta}_{\text{in}}$, into the basis waveform, by taking the Fourier transform of Equation (2.8) of [45]:

$$X_i(\omega;\boldsymbol{\theta}) = e^{-i\omega(t_\oplus - \mathbf{d}_i\cdot\mathbf{n})}\frac{r_{1,i}}{r}e^{2i\phi_c} \left[\frac{1}{2}\left(1+\cos^2\iota\right)\Re\{\zeta\} - i\left(\cos\iota\right)\Im\{\zeta\}\right] H(\omega;\boldsymbol{\theta}_{\text{in}}) \quad (6)$$

for $\omega \geq 0$, where

$$\zeta = e^{-2i\psi}\left(F_{+,i}(\alpha,\delta,t_\oplus) + iF_{\times,i}(\alpha,\delta,t_\oplus)\right). \quad (7)$$

The quantities $F_{+,i}$ and $F_{\times,i}$ are the dimensionless detector antenna factors, defined such that $0 \leq {F_{+,i}}^2 + {F_{\times,i}}^2 \leq 1$. They depend on the orientation of detector i as well as the sky location and sidereal time of the event and are presented in [46]. In a coordinate system with the x and y axes aligned with the arms of a detector, the antenna pattern is given in spherical polar coordinates as

$$F_+ = -\frac{1}{2}(1+\cos^2\theta)\cos 2\phi, \quad (8)$$

$$F_\times = -\cos\theta \sin 2\phi. \quad (9)$$

The unit vector $\mathbf{d}_i$ represents the position of detector i in units of light travel time.[5] The vector $\mathbf{n}$ is the direction of the source. The negative sign in the dot product $-\mathbf{d}_i\cdot\mathbf{n}$ is present because the direction of travel of the GW signal is opposite to that of its sky location. The quantity $r_{1,i}$ is a fiducial distance at which detector i would register SNR=1 for an optimally oriented binary (face-on, and in a direction perpendicular to the interferometer's arms):

$$r_{1,i} = 1/\sigma_i, \qquad {\sigma_i}^2 = \int_0^\infty \frac{|H(\omega;\boldsymbol{\theta}_{\text{in}})|^2}{S_i(\omega)}\, d\omega. \quad (10)$$

More succinctly, we can write the signal received by detector i in terms of observable extrinsic parameters $\boldsymbol{\theta}_i = (\rho_i, \gamma_i, \tau_i)$, the amplitude ρ_i, phase γ_i, and time delay τ_i on arrival at detector i:

$$X_i(\omega;\boldsymbol{\theta}_i,\boldsymbol{\theta}_{\text{in}}) = X_i(\omega;\rho_i,\gamma_i,\tau_i,\boldsymbol{\theta}_{\text{in}}) = \frac{\rho_i}{\sigma_i}e^{i(\gamma_i - \omega\tau_i)}H(\omega;\boldsymbol{\theta}_{\text{in}}). \quad (11)$$

The prevailing technique for detection of GWs from CBCs is to realize a maximum likelihood (ML) estimator (MLE) from the likelihood in Equation (3) and the

[4] This list of parameters involves some simplifying assumptions. Eccentricity is omitted: although it may play a major role in the evolution and waveforms of rare close binaries formed by dynamical capture [39–41], BNS systems formed by binary stellar evolution should almost always circularize due to tidal interaction [42] and later GW emission [43] long before the inspiral enters LIGO's frequency range of ∼10–1000 kHz. Tidal deformabilities of the NSs are omitted because the signal imprinted by the companions' material properties is so small that it will only be detectable by an Einstein Telescope-class GW observatory [44]. Furthermore, in GW detection efforts, especially those focused on BNS systems, the component spins $\mathbf{S}_1$ and $\mathbf{S}_2$ are often assumed to be nonprecessing and aligned with the system's total angular momentum and condensed to a single scalar parameter χ, or even neglected entirely: $\mathbf{S}_1 = \mathbf{S}_2 = 0$.

[5] When considering transient GW sources such as those that we are concerned with in this thesis, the origin of the coordinate system is usually taken to be the geocenter. For long-duration signals such as those from statically deformed neutron stars, the solar system barycenter is a more natural choice.

signal model in Equation (11). Concretely, this results in a bank of matched filters, or cross-correlations between the incoming data stream and a collection of template waveforms,

$$z_i(\tau_i;\boldsymbol{\theta}_{\rm in}) = \frac{1}{\sigma_i(\boldsymbol{\theta}_{\rm in})}\int_0^\infty \frac{H^*(\omega;\boldsymbol{\theta}_{\rm in})Y_i(\omega)e^{i\omega\tau_i}}{S_i(\omega)}\,d\omega. \quad (12)$$

The ML point estimates of the signal parameters, $\mathrm{MLE}(\mathbf{y}) = \{\{\hat{\boldsymbol{\theta}}_i\}_i, \hat{\boldsymbol{\theta}}_{\rm in}\} = \{\{\hat\rho_i,\hat\gamma_i,\hat\tau_i\}_i, \hat{\boldsymbol{\theta}}_{\rm in}\}$, are given by

$$\hat{\boldsymbol{\theta}}_{\rm in}, \{\hat\tau_i\}_i = \underset{\boldsymbol{\theta}_{\rm in},\{\hat\tau_i\}_i}{\operatorname{argmax}} \sum_i \left|z_i\left(\tau_i;\boldsymbol{\theta}_{\rm in}\right)\right|^2, \quad (13)$$

$$\hat\rho_i = \left|z_i\left(\hat\tau_i;\hat{\boldsymbol{\theta}}_{\rm in}\right)\right|, \quad (14)$$

$$\hat\gamma_i = \arg z_i\left(\hat\tau_i;\hat{\boldsymbol{\theta}}_{\rm in}\right). \quad (15)$$

A detection candidate consists of $\{\{\hat\rho_i,\hat\gamma_i,\hat\tau_i\}_i, \hat{\boldsymbol{\theta}}_{\rm in}\}$. There are various ways to characterize the significance of a detection candidate. In Gaussian noise, the maximum likelihood for the network is obtained by maximizing the network S/N, $\rho_{\rm net}$,

$$\hat\rho_{\rm net} = \max_{\boldsymbol{\theta}} \sum_i |z_i(\boldsymbol{\theta})|^2 = \sqrt{\sum_i \hat\rho_i^2}; \quad (16)$$

this, therefore, is the simplest useful candidate ranking statistic.

A. Uncertainty and the Fisher matrix

We can predict the uncertainty in the detection pipeline's ML estimates using the CRLB. The CRLB has been widely applied in GW data analysis to estimate parameter estimation uncertainty [3, 17, 18, 47–49][6]. As we noted, there are significant caveats to the CRLB at low or moderate S/N [35–38]. However, here we will be concerned more with gaining intuition from the block structure of the Fisher matrix than its numerical value. Furthermore, the Fisher matrix in its own right—independent of its suitability to describe the parameter covariance—is a well-defined property of any likelihood function, and we will exploit it as such in Section III.

We will momentarily consider the likelihood for a single detector:

$$\mathcal{L}\left(Y_i;\rho_i,\gamma_i,\tau_i,\boldsymbol{\theta}_{\rm in}\right) \propto \exp\left[-\frac{1}{2}\int_0^\infty \frac{\left|Y_i(\omega) - X_i\left(\omega;\rho_i,\gamma_i,\tau_i,\boldsymbol{\theta}_{\rm in}\right)\right|^2}{S_i(\omega)}\,d\omega\right], \quad (17)$$

with $X_i(\omega;\rho_i,\gamma_i,\tau_i,\boldsymbol{\theta}_{\rm in})$ given by Equation (11).

The Fisher information matrix for a measurement y described by the unknown parameter vector $\boldsymbol{\theta}$ is the conditional expectation value

$$\mathcal{I}_{jk} = \mathrm{E}\left[\left(\frac{\partial\log\mathcal{L}(Y_i;\boldsymbol{\theta})}{\partial\theta_j}\right)\left(\frac{\partial\log\mathcal{L}(Y_i;\boldsymbol{\theta})}{\partial\theta_k}\right)\middle|\boldsymbol{\theta}\right]. \quad (18)$$

The Fisher matrix describes how strongly the likelihood depends, on average, on the parameters. Furthermore, it provides an estimate of the mean-square error in the parameters. If $\hat{\boldsymbol{\theta}}$ is an unbiased estimator of $\boldsymbol{\theta}$, $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ is the measurement error, and $\Sigma = \mathrm{E}[\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^{\mathsf{T}}]$ is the covariance of the measurement error, then the CRLB says that $\Sigma \geq \mathcal{I}^{-1}$, in the sense that $\left(\Sigma - \mathcal{I}^{-1}\right)$ is positive semi-definite.

Note that if $\log\mathcal{L}$ is twice differentiable in terms of $\boldsymbol{\theta}$, then the Fisher matrix can also be written in terms of second derivatives as

$$\mathcal{I}_{jk} = \mathrm{E}\left[-\frac{\partial^2\log\mathcal{L}(Y_i;\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\middle|\boldsymbol{\theta}\right]. \quad (19)$$

When (as in our assumptions) the likelihood is Gaussian,[7] Equation (18) simplifies to

$$\mathcal{I}_{jk} = \int_0^\infty \Re\left[\left(\frac{\partial X_i}{\partial\theta_j}\right)^*\left(\frac{\partial X_i}{\partial\theta_k}\right)\right]\frac{1}{S_i(\omega)}\,d\omega. \quad (20)$$

This form is useful because it involves manipulating the signal $X_i(\omega)$ rather than the entire observation $Y(\omega)$. In terms of the kth S/N-weighted moment of angular frequency,

$$\overline{\omega^k}_i = \left[\int_0^\infty \frac{|h(\omega)|^2}{S_i(\omega)}\omega^k\,d\omega\right]\left[\int_0^\infty \frac{|h(\omega)|^2}{S_i(\omega)}\,d\omega\right]^{-1}, \quad (21)$$

the Fisher matrix for the signal in the ith detector is

$$\mathcal{I}_i = \begin{pmatrix} \mathcal{I}_{\boldsymbol{\theta}_i,\boldsymbol{\theta}_i} & \mathcal{I}_{\boldsymbol{\theta}_i,\boldsymbol{\theta}_{\rm in}} \\ \mathcal{I}^{\mathsf{T}}_{\boldsymbol{\theta}_i,\boldsymbol{\theta}_{\rm in}} & {\rho_i}^2\mathcal{I}_{\boldsymbol{\theta}_{\rm in},\boldsymbol{\theta}_{\rm in}} \end{pmatrix}. \quad (22)$$

The top-left block describes only the extrinsic parameters, and is given by

$$\mathcal{I}_{\boldsymbol{\theta}_i,\boldsymbol{\theta}_i} = \begin{array}{c} \\ \rho_i \\ \gamma_i \\ \tau_i \end{array}\begin{array}{c} \begin{array}{ccc} \rho_i & \gamma_i & \tau_i \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & {\rho_i}^2 & -{\rho_i}^2\overline{\omega}_i \\ 0 & -{\rho_i}^2\overline{\omega}_i & {\rho_i}^2\overline{\omega^2}_i \end{pmatrix} \end{array}. \quad (23)$$

[6] The Fisher matrix is also used in construction of CBC matched filter banks. The common procedure is to place templates uniformly according to the determinant of the signal space metric, which is the Fisher matrix. This is equivalent to uniformly sampling the Jeffreys prior. In practice, this is done either by constructing a hexagonal lattice [50] or sampling stochastically [51–55].

[7] This assumes that the merger occurs at a frequency outside the sensitive "bucket" of the detector's noise PSD. There are additional terms if the GW spectrum drops to zero within the sensitive bandwidth of the detector, as can be the case for neutron star–black hole (NSBH) mergers; see [56].

(This is equivalent to an expression given in [25].) The bottom row and right column of Equation (22) describe the intrinsic parameters and how they are coupled to the extrinsic parameters. We show in Appendix A that we need not consider the intrinsic parameters at all if we are concerned only with sky localization.

For our likelihood, the CRLB implies that

$$\operatorname{cov}\begin{pmatrix}\tilde{\rho}_i\\ \tilde{\gamma}_i\\ \tilde{\tau}_i\end{pmatrix} \geq \mathcal{I}^{-1} = \begin{pmatrix}1 & 0 & 0\\ 0 & {\rho_i}^2\overline{\omega^2}_i/{\omega_{\mathrm{rms},i}}^2 & {\rho_i}^2\overline{\omega}_i/{\omega_{\mathrm{rms},i}}^2\\ 0 & {\rho_i}^2\overline{\omega}_i/{\omega_{\mathrm{rms},i}}^2 & {\rho_i}^2/{\omega_{\mathrm{rms},i}}^2\end{pmatrix}, \tag{24}$$

where ${\omega_{\mathrm{rms},i}}^2 = \overline{\omega^2}_i - {\overline{\omega}_i}^2$. This structure implies that errors in S/N are uncorrelated with errors in time and phase, and that there is a particular sum and difference of the times and phases that are measured independently (see Appendix B).

Reading off the $\tau\tau$ element of the covariance matrix reproduces the timing accuracy in Equation (24) of [17],

$$\operatorname{std}\left(\hat{\tau}_i - \tau_i\right) \geq \sqrt{\left(\mathcal{I}^{-1}\right)_{\tau\tau}} = \frac{\rho_i}{\omega_{\mathrm{rms},i}}. \tag{25}$$

Fairhurst [17] goes on to frame the characteristic position reconstruction accuracy of a GW detector network in terms of time delay triangulation, with the above formula describing the time of arrival uncertainty for each detector. In Appendix C, we show how to extend this formalism to include the phases and amplitudes on arrival as well.

II. BAYESIAN PROBABILITY AND PARAMETER ESTIMATION

In the Bayesian framework, parameters are inferred from the data by forming the posterior distribution, $p(\boldsymbol{\theta}|\mathbf{y})$, which describes the probability of the parameters given the observations. Bayes' rule relates the likelihood $p(\mathbf{y}|\boldsymbol{\theta})$ to the posterior $p(\boldsymbol{\theta}|\mathbf{y})$,

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}, \tag{26}$$

introducing the prior distribution $p(\boldsymbol{\theta})$ which encapsulates previous information about the paramters (for example, arising from earlier observations or from known physical bounds on the parameters), and the evidence $p(\mathbf{y})$ which can be thought of as a normalization factor or as describing the parsimoniousness of the model.

The choice of prior is determined by one's astrophysical assumptions. During LIGO's sixth science run (S6) when LIGO's Bayesian CBC parameter estimation pipelines were first deployed, the prior was taken to be isotropic in sky location and binary orientation, and uniform in volume, arrival time, and the component masses [31].

In Bayesian inference, although it is often easy to write down the likelihood or even the full posterior in closed form, usually one is interested in only a subset $\boldsymbol{\beta}$ of all of the model's parameters, the others $\boldsymbol{\lambda}$ being nuisance parameters. In this case, we integrate away the nuisance parameters, forming the marginal posterior

$$p(\boldsymbol{\beta}|\mathbf{y}) = \int \frac{p(\mathbf{y}|\boldsymbol{\beta},\boldsymbol{\lambda})p(\boldsymbol{\beta},\boldsymbol{\lambda})}{p(\mathbf{y})}\, d\boldsymbol{\lambda} \tag{27}$$

with $\boldsymbol{\theta} = (\boldsymbol{\beta},\boldsymbol{\lambda})$. For instance, for the purpose of locating a GW source on the sky, all parameters but (α,δ) are nuisance parameters.

III. THE BAYESTAR LIKELIHOOD

For the purpose of rapid sky localization, we assume that we do not have access to the GW data $\mathbf{Y}$ itself, and that our only contact with it is through the ML parameter estimates $\{\{\hat{\rho}_i,\hat{\gamma}_i,\hat{\tau}_i\}_i, \hat{\boldsymbol{\theta}}_{\mathrm{in}}\}$. Although this is a significant departure from conventional GW parameter estimation techniques, we can still apply the full Bayesian machinery of Equation (27) to compute a posterior distribution for the sky location.

The relevant likelihood is now the probability of the ML estimates, conditioned upon the true parameter values, and marginalized over all possible GW observations:

$$p\left(\{\hat{\boldsymbol{\theta}}_i\}_i, \hat{\boldsymbol{\theta}}_{\mathrm{in}}\middle|\boldsymbol{\theta}\right) \propto \int_{\mathbf{Y}|\mathrm{MLE}(\mathbf{Y})=\{\{\hat{\boldsymbol{\theta}}_i\}_i,\hat{\boldsymbol{\theta}}_{\mathrm{in}}\}} p(\mathbf{Y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\mathbf{Y}. \tag{28}$$

Although we may not be able to evaluate this equation directly, with some educated guesses we can create a likelihood that has many properties in common with it. Any valid approximate likelihood must have the same Fisher matrix as shown in Equation (22). It must also have the same limiting behavior: it should be periodic in the phase error $\tilde{\gamma}_i$ and go to zero as $\tilde{\tau}_i \to \pm\infty$, $\hat{\rho}_i \to 0$, or $\hat{\rho}_i \to \infty$. Additionally, when $\tilde{\tau}_i = 0$, the distribution of $\hat{\rho}_i^2$ should reduce to a noncentral χ^2 distribution with two degrees of freedom, centered about ${\rho_i}^2$, because the complex matched filter time series $z_i(t)$ is Gaussian (under the ideal assumption the GW strain time series is Gaussian).

These conditions could be satisfied by realizing a multivariate Gaussian distribution with covariance matrix $\Sigma = \mathcal{I}^{\intercal}$, and then replacing individual quadratic terms in the exponent of the form $-\tilde{\theta}^2/2$ with $\cos\tilde{\theta}$.

A more natural way is to plug the signal model from Equation (11) *evaluated at the ML parameter estimates* into the single-detector likelihood in Equation (17):

$$p\left(\hat{\boldsymbol{\theta}}_i\middle|\boldsymbol{\theta}\right) := p\left(Y_i(\omega) = X_i(\omega;\hat{\boldsymbol{\theta}})\middle|\boldsymbol{\theta}\right) \propto \exp\left[-\frac{1}{2}\int_0^\infty \left|\frac{\hat{\rho}_i}{\sigma_i(\hat{\boldsymbol{\theta}}_{\rm in})}e^{i(\hat{\gamma}_i-\omega\hat{\tau}_i)}\frac{H(\omega;\hat{\boldsymbol{\theta}}_{\rm in})}{S_i(\omega)} - \frac{\rho_i}{\sigma_i(\boldsymbol{\theta}_{\rm in})}e^{i(\gamma_i-\omega\tau_i)}\frac{H(\omega;\boldsymbol{\theta}_{\rm in})}{S_i(\omega)}\right|^2 d\omega\right]. \tag{29}$$

If we further assume that *the intrinsic parameters are equal to their ML estimates*, $\boldsymbol{\theta}_{\rm in} = \hat{\boldsymbol{\theta}}_{\rm in}$, then this reduces to what we call the *autocorrelation likelihood*,

$$p\left(\hat{\rho}_i, \hat{\gamma}_i, \hat{\tau}_i|\rho_i, \gamma_i, \tau_i\right) \propto \exp\left[-\frac{1}{2}\hat{\rho}_i^2 - \frac{1}{2}{\rho_i}^2 + \hat{\rho}_i\rho_i\Re\left\{e^{i\tilde{\gamma}_i}a_i^*(\tilde{\tau}_i)\right\}\right], \tag{30}$$

with $\tilde{\gamma}_i = \hat{\gamma}_i - \gamma_i$, $\tilde{\tau}_i = \hat{\tau}_i - \tau_i$, and the template's autocorrelation function $a_i(t;\boldsymbol{\theta}_{\rm in})$ defined as

$$a_i(t;\boldsymbol{\theta}_{\rm in}) := \frac{1}{{\sigma_i}^2(\boldsymbol{\theta}_{\rm in})}\int_0^\infty \frac{\left|H(\omega;\hat{\boldsymbol{\theta}}_{\rm in})\right|^2}{S_i(\omega)}e^{i\omega t}\,d\omega. \tag{31}$$

Some example autocorrelation functions and corresponding likelihoods are shown in Figure 1. To assemble the joint likelihood for the whole network, we form the product of the autocorrelation likelihoods from the individual detectors:

$$p\left(\{\hat{\rho}_i, \hat{\gamma}_i, \hat{\tau}_i\}_i|\{\rho_i, \gamma_i, \tau_i\}_i\right) \propto \exp\left[-\frac{1}{2}\sum_i\hat{\rho}_i^2 - \frac{1}{2}\sum_i{\rho_i}^2 + \sum_i\hat{\rho}_i\rho_i\Re\left\{e^{i\tilde{\gamma}_i}a^*(\tilde{\tau}_i)\right\}\right]. \tag{32}$$

In the following section, we discuss some key properties of the autocorreltion likelihood.

A. Properties

First, the autocorrelation likelihood has the elegant feature that if we were to replace the autocorrelation function with the S/N time series for the best-matching template, $z(\tau;\hat{\theta}_{\rm in})$, we would recover the likelihood for the full GW time series, evaluated at the ML estimate of the intrinsic parameters, *viz.*:

$$\exp\left[-\frac{1}{2}\sum_i{\rho_i}^2 + \sum_i\rho_i\Re\left\{e^{-i\gamma_i}z_i^*(\tau_i)\right\}\right]. \tag{33}$$

(We have ommitted the term $\int|Y_i(\omega)|^2/S(\omega)d\omega$, which takes the place of the earlier $\hat{\rho}_i^2$ term and is only important for normalization.) The numerical scheme that we will develop is thus equally applicable for rapid, coincidence-based localization, or as a fast extrinsic marginalization step for the full parameter estimation.

Second, observe that at the true parameter values, $\hat{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_i$, the logarithms of Equation (30) and Equation (17) have the same Jacobian. This is because the derivatives of the autocorrelation function are

$$a^{(n)}(t) = i^n\overline{\omega^n},$$

with $\overline{\omega^n}$ defined in Equation (21). For example, the first few derivatives are

$$a(0) = 1, \qquad \dot{a}(0) = i\overline{\omega}, \qquad \ddot{a}(0) = -\overline{\omega^2}.$$

Using Equation (19), we can compute the Fisher matrix elements for the autocorrelation likelihood given by Equation (30), with detector subscript suppressed:

$$\begin{aligned}
\mathcal{I}_{\rho\rho} &= 1,\\
\mathcal{I}_{\rho\gamma} &= 0,\\
\mathcal{I}_{\rho\tau} &= 0,\\
\mathcal{I}_{\gamma\gamma} &= \rho^2\int_{-T}^{T}|a(t)|^2\,w(t;\rho)dt, &(34)\\
\mathcal{I}_{\tau\tau} &= -\rho^2\int_{-T}^{T}\Re\left[a^*(t)\ddot{a}(t)\right]w(t;\rho)dt, &(35)\\
\mathcal{I}_{\gamma\tau} &= -\rho^2\int_{-T}^{T}\Im\left[a^*(t)\dot{a}(t)\right]w(t;\rho)dt, &(36)
\end{aligned}$$

where

$$w(t;\rho) = \frac{\exp\left[\frac{\rho^2}{4}|a(t)|^2\right]\left(I_0\left[\frac{\rho^2}{4}|a(t)|^2\right] + I_1\left[\frac{\rho^2}{4}|a(t)|^2\right]\right)}{2\int_{-T}^{T}\exp\left[\frac{\rho^2}{4}|a(t')|^2\right]I_0\left[\frac{\rho^2}{4}|a(t')|^2\right]dt'}. \tag{37}$$

The notation I_k denotes a modified Bessel function of the first kind. Matrix elements that are not listed have values that are implied by the symmetry of the Fisher matrix. Note that the minus signs are correct but a little confusing: despite them, $\mathcal{I}_{\gamma\gamma}, \mathcal{I}_{\tau\tau} \geq 0$ and $\mathcal{I}_{\gamma\tau} \leq 0$. The time integration limits $[-T, T]$ correspond to a flat prior on arrival time, or a time coincidence window between detectors.

We can show that the weighting function $w(t;\rho)$ approaches a Dirac delta function as $\rho \to \infty$, so that

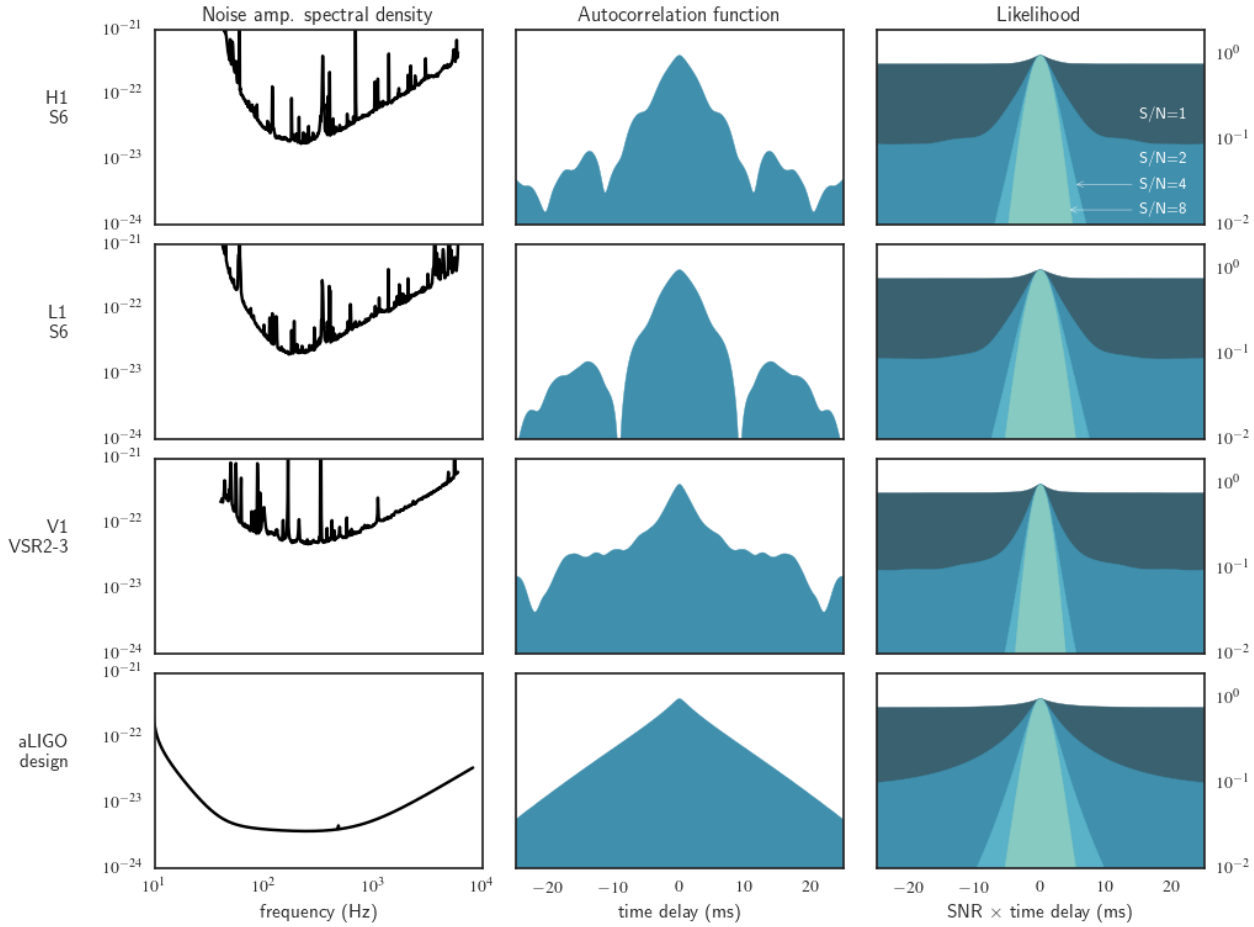

FIG. 1. The autocorrelation likelihood for a $(1.4, 1.4)\,M_\odot$ binary as observed by four detector configurations: from top to bottom, the final sensitivity achieved by the LIGO Hanford, LIGO Livingston, and Virgo detectors in their "initial" configuration, and the final Advanced LIGO design sensitivity. The left panels show the noise amplitude spectral densities. The middle panels show the absolute value of the autocorrelation function. The right panel shows the phase-marginalized autocorrelation likelihood for S/N=1, 2, 4, and 8. In the right panel, the time scale is normalized by S/N so that one can see that as S/N increases, a central parabola is approached (the logarithm of a Gaussian distribution with standard deviation given by the Fisher matrix).

the Fisher matrix for the autocorrelation likelihood approaches the Fisher matrix for the full GW data, Equation (23), as $\rho \to \infty$. The Bessel functions asymptotically approach:

$$I_0(x), I_1(x) \to \frac{e^x}{\sqrt{2\pi x}} \text{ as } x \to \infty.$$

For large ρ, the exponents of e^{ρ^2} dominate Equation (37) and we can write:

$$w(t;\rho) \to \frac{\exp\left[\frac{\rho^2}{2}|a(t)|^2\right]}{\int_{-T}^{T} \exp\left[\frac{\rho^2}{2}|a(t')|^2\right] dt'} \text{ as } \rho \to \infty.$$

The Taylor expansion of $|a(t)|^2$ is

$$\begin{aligned} |a(t)|^2 &= 1 + \frac{1}{2}\left(\frac{\partial^2}{\partial t^2}|a(t)|^2\Big|_{t=0}\right) t^2 + \mathcal{O}(t^4) \\ &= 1 - {\omega_{\rm rms}}^2 t^2 + \mathcal{O}(t^4). \end{aligned}$$

Substituting, we find that $w(t;\rho)$ approaches a normalized Gaussian distribution:

$$w(t;\rho) \approx \frac{\exp\left[-\frac{1}{2}\rho^2 {\omega_{\rm rms}}^2 t^2\right]}{\int_{-T}^{T} \exp\left[-\frac{1}{2}\rho^2 {\omega_{\rm rms}}^2 (t')^2\right] dt'}.$$

And finally, because the Dirac delta function may be de-

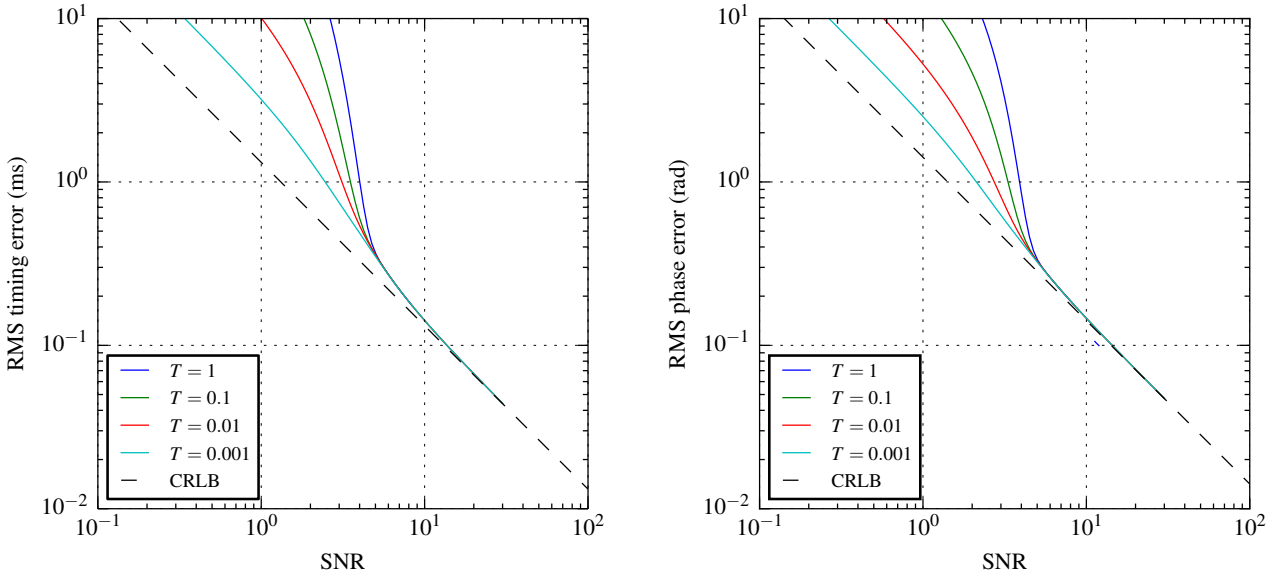

FIG. 2. CRLB on RMS timing uncertainty and phase error, using the likelihood for the full GW data (Equation 17; dashed diagonal line) or the autocorrelation likelihood (Equation 30; solid lines) with a selection of arrival time priors.

fined as the limit of a Gaussian, $w(t;\rho) \to \delta(t)$ as $\rho \to \infty$.

We can now write the Fisher matrix for the autocorrelation likelihood in a way that makes a comparison to the full signal model explicit. Define:

$$\mathcal{I}_{\gamma\gamma} = \rho^2 \cdot \mathfrak{Q}_{\gamma\gamma}(\rho),$$
$$\mathcal{I}_{\tau\tau} = \rho^2\overline{\omega^2} \cdot \mathfrak{Q}_{\tau\tau}(\rho),$$
$$\mathcal{I}_{\gamma\tau} = -\rho^2\overline{\omega} \cdot \mathfrak{Q}_{\gamma\tau}(\rho).$$

Now, the $\mathfrak{Q}_{ij}$[8] contain the integrals from Equations (34, 35, 36) and encode the departure of the autocorrelation likelihood from the likelihood of the full data at low S/N. All of the $\mathfrak{Q}_{ij}(\rho)$ are sigmoid-type functions that asymptotically approach 1 as $\rho \to \infty$ (see Figures 2 and 3). The transition S/N $\rho_{\rm crit}$ is largely the same for all three non-trivial matrix elements, and is determined by the time coincidence window T and the signal bandwidth $\omega_{\rm rms}$.

In the limit of large S/N, our interpretation is that the point estimates $(\hat{\rho}, \hat{\gamma}, \hat{\tau})$ contain all of the information about the underlying extrinsic parameters.

On the other hand, in the low S/N limit, the diminishing value of $\mathfrak{Q}_{ij}(\rho)$ reflects the fact that some information is lost when the full data $\mathbf{x}$ is discarded. Concretely, as the prior interval T becomes large compared to $1/\rho\omega_{\rm rms}$, the ML estimator becomes more and more prone to picking up spurious noise fluctuations far from the true signal. Clearly, when the coincidence window T is kept small as possible, more information is retained in the ML point estimates. Put another way, if T is small, then the transition S/N $\rho_{\rm crit}$ is also small and fainter signals become useful for parameter estimation. In this way, the BAYESTAR likelihood exhibits the *threshold effect* that is well-known in communication and radar applications [57–59].

In the following sections, we describe our prior and our numerical scheme to integrate over nuisance parameters, which amounts to the BAYESTAR algorithm.

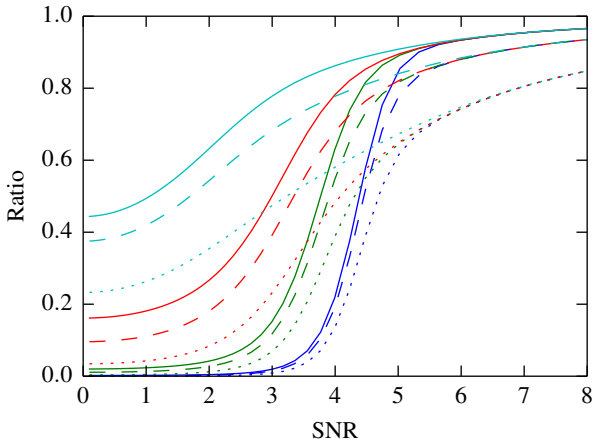

FIG. 3. Ratio between Fisher matrix elements (solid: $\mathfrak{Q}_{\gamma\gamma}$, dashed: $\mathfrak{Q}_{\gamma\tau}$, dotted: $\mathfrak{Q}_{\tau\tau}$) for the autocorrelation likelihood and the full GW data. Colors correspond to different arrival time priors as in Figure 2.

[8] The Fish(er) factor.

IV. PRIOR AND PROBLEM SETUP

The detection pipeline supplies a candidate, $\{\{\hat{\rho}_i, \hat{\gamma}_i, \hat{\tau}_i\}_i, \hat{\boldsymbol{\theta}}_{\rm in}\}$, and discretely sampled noise PSDs, $S_i(\omega_j)$, for all detectors. We compute the GW signal for a source with intrinsic parameters equal to the detection pipeline's estimate, $H(\omega; \hat{\boldsymbol{\theta}}_{\rm in})$. Then we find the S/N=1 horizon distance $r_{1,i}$ for each detector by numerically integrating Equation (10).

We have no explicit prior on the intrinsic parameters; in our analysis they are fixed at their ML estimates, $\hat{\boldsymbol{\theta}}_{\rm in}$.[9]

The arrival time prior is connected to the origin of the detector coordinate system. Given the Earth-fixed coordinates of the detectors $\mathbf{n}_i$ and the arrival times τ_i, we compute their averages weighted by the timing uncertainty formula:

$$\langle \mathbf{n} \rangle = \frac{\sum_i \dfrac{\mathbf{n}_i}{(\hat{\rho}_i \omega_{{\rm rms},i})^2}}{\sum_i \dfrac{1}{(\hat{\rho}_i \omega_{{\rm rms},i})^2}}, \qquad \langle \hat{\tau} \rangle = \frac{\sum_i \dfrac{\hat{\tau}_i}{(\hat{\rho}_i \omega_{{\rm rms},i})^2}}{\sum_i \dfrac{1}{(\hat{\rho}_i \omega_{{\rm rms},i})^2}}.$$

Then, we subtract these means:

$$\mathbf{n}_i \leftarrow \mathbf{n}_i - \langle \mathbf{n} \rangle, \qquad \hat{\tau}_i \leftarrow \hat{\tau}_i - \langle \hat{\tau} \rangle.$$

In these coordinates, now relative to the weighted detector array barycenter, the arrival time prior is uniform in $-T \le t \le T$, with $T = \max_i |\mathbf{n}_i|/c + 5$ ms.

The distance prior is a user-selected power of distance,

$$p(r) \propto \begin{cases} r^m & \text{if } r_{\rm min} < r < r_{\rm max} \\ 0 & \text{otherwise,} \end{cases}$$

where $m = 2$ for a prior that is uniform in volume, and $m = -1$ for a prior that is uniform in the logarithm of the distance. If a distance prior is not specified, the default is uniform in volume out to the maximum SNR=4 horizon distance:

$$m = 2, \qquad r_{\rm min} = 0, \qquad r_{\rm max} = \frac{1}{4} \max_i r_{1,i}.$$

Finally, the prior is uniform in $-1 \le \cos\iota \le 1$ and $0 \le \psi < \pi$.

We compute the autocorrelation function for each detector from $t = 0$ to $t = T$ at intervals of $\Delta t = 1/f_s$, where f_s is the smallest power of two that is greater than or equal to the Nyquist rate. Because BNS signals typically terminate at about 1500 Hz, a typical value for Δt is $(4096\,{\rm Hz})^{-1}$. We use a pruned fast Fourier transform (FFT) because for BNS systems, the GW signal remains in LIGO's sensitive band for ∼100–1000 s, whereas $T \sim 10$ ms.[10]

[9] As noted in footnote 6, the detection template bank is typically designed to uniformly sample the Jeffreys prior on the intrinsic parameters. Due to the equivalence of marginalization and maximization with respect to a parameter under a Gaussian distribution, fixing the intrinsic parameters at their ML estimates is roughly equivalent to selecting the Jeffreys prior.

[10] See http://www.fftw.org/pruned.html for a discussion of methods for computing the pruned FFT, the first K samples of an FFT of length N.

V. MARGINAL POSTERIOR

The marginal posterior as a function of sky location is

$$f(\alpha,\delta) \propto \int_0^{\pi} \int_{-1}^{1} \int_{-T}^{T} \int_{r_{\rm min}}^{r_{\rm max}} \int_0^{2\pi} \exp\left[-\frac{1}{2}\sum_i {\rho_i}^2 + \sum_i \hat{\rho}_i \rho_i \Re\left\{e^{i\tilde{\gamma}_i} a^*(\tilde{\tau}_i)\right\}\right] r^m d\phi_c\, dr\, dt_\oplus\, d\cos\iota\, d\psi. \quad (38)$$

To marginalize over the coalescence phase, we can write $\tilde{\gamma}_i = \tilde{\gamma}_i' + 2\phi_c$. Then integrating over ϕ_c and suppressing normalization factors, we get

$$f(\alpha,\delta) \to \int_0^{\pi} \int_{-1}^{1} \int_{-T}^{T} \int_{r_{\rm min}}^{r_{\rm max}} \exp\left[-\frac{1}{2}\sum_i {\rho_i}^2\right] I_0\left[\left|\sum_i \hat{\rho}_i \rho_i e^{i\tilde{\gamma}_i} a_i^*(\tilde{\tau}_i)\right|\right] r^m dr\, dt_\oplus\, d\cos\iota\, d\psi. \quad (39)$$

In the above equation, we need not distinguish between $\tilde{\gamma}_i$ and $\tilde{\gamma}_i'$ because the likelihood is now invariant under arbitrary phase shifts of all of the detectors' signals.

A. Integral over angles and time

The integrand is periodic in ψ, so simple Newton–Cotes quadrature over ψ exhibits extremely rapid convergence (see Figure 4). We therefore sample the posterior on a regular grid of 10 points from 0 to π.

The integral over $\cos\iota$ converges just as rapidly with Gauss–Legendre quadrature (see Figure 4), so we use a 10-point Gauss–Legendre rule for integration over $\cos\iota$.

We sample $t_\oplus$ regularly from $-T$ to T at intervals of Δt. This is typically $\sim 2(10\,{\rm ms})(4096\,{\rm Hz}) \approx 80$ samples. We use Catmull–Rom cubic splines to interpolate the real and imaginary parts of the autocorrelation functions between samples.

B. Integral over distance

The distance integral is now performed differently from what we initially described in [60, 61]; the method de-

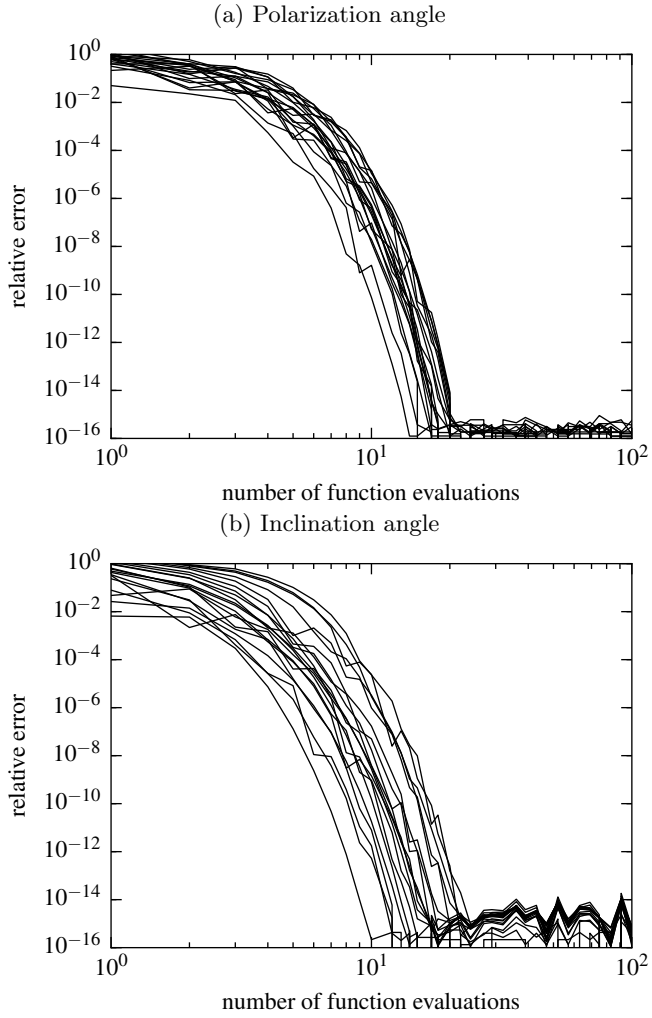

FIG. 4. Relative error in BAYESTAR integration scheme as a function of the number of Gaussian quadrature nodes. The two panels describe (a) the integral over the polarization angle ψ and (b) the integral over inclination angle ι.

scribed in the present work is about an order of magnitude faster. We define $\rho_i = \omega_i/r$ in order to absorb all of the distance-independent terms in the amplitudes into ω_i, and then define

$$p^2 = \frac{1}{2}\sum_i {\omega_i}^2 \tag{40}$$

$$b = \left|\sum_i \hat{\rho}_i \omega_i e^{i\tilde{\gamma}_i} a_i^*(\tilde{\tau}_i)\right|. \tag{41}$$

The innermost integral over distance r may then be written as

$$\begin{aligned}\mathscr{F} &= \int_{r_{\min}}^{r_{\max}} \exp\left[-\frac{p^2}{r^2}\right] I_0\left[\frac{b}{r}\right] r^m dr \\ &= \int_{r_{\min}}^{r_{\max}} \exp\left[-\frac{p^2}{r^2}+\frac{b}{r}\right] \overline{I}_0\left[\frac{b}{r}\right] r^m dr\end{aligned} \tag{42}$$

or, completing the square,

$$\mathscr{F} = \exp\left[\frac{p^2}{{r_0}^2}\right] \int_{r_{\min}}^{r_{\max}} \exp\left[-\left(\frac{p}{r}-\frac{p}{r_0}\right)^2\right] \overline{I}_0\left[\frac{2p^2}{r r_0}\right] r^m dr \tag{43}$$

$$= \exp\left[\frac{p^2}{{r_0}^2}\right] \mathscr{G}, \tag{44}$$

where

$$r_0 = 2p^2/b \tag{45}$$

$$\overline{I}_0(x) = \exp(-|x|) I_0(x). \tag{46}$$

The coefficients p^2 and b are nonnegative and independent of distance. p has a maximum value of

$$p_{\max} = \sqrt{\frac{1}{2}\sum_i \left(\frac{r_{1,i}}{r}\right)^2}. \tag{47}$$

The symbol $\overline{I}_0$ denotes an exponentially scaled Bessel function. In the limit of large argument, $I_0(|x|) \sim$

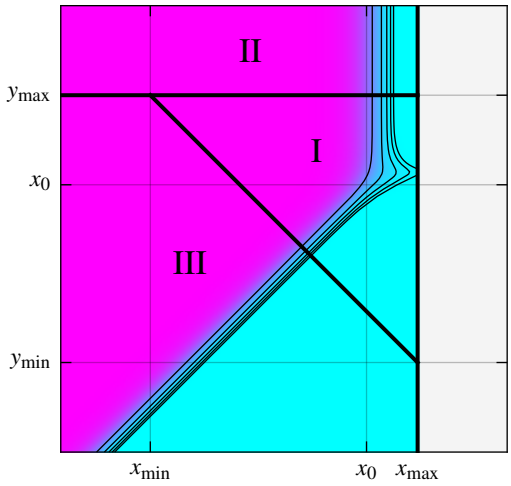

FIG. 5. Partition of the parameter space of the distance integral into three regions for (bi)cubic interpolation.

$\exp(|x|)/\sqrt{2\pi|x|}$ [62, 63][11]. The scaled Bessel function is useful for evaluation on a computer because it has a relatively small range $(0, 1]$ and varies slowly in proportion to $x^{1/2}$.

1. *Parameter grid*

This integral is not particularly amenable to low-order Gaussian quadrature. However, luckily $\mathscr{G}$ is a very well-behaved function of p and r_0, so we evaluate it using a lookup table and bicubic interpolation. The lookup table is produced in logarithmic coordinates

$$x = \log p, \qquad y = \log r_0. \tag{48}$$

As shown in Figure 5, the function basically consists of a plateau region in the upper-left half of the plane delimited by the lines $y = x$ and $x = \log p_0$, with

$$p_0 = \frac{1}{2}\begin{cases} r_{\max} & \text{if } m \geq 0 \\ r_{\min} & \text{if } m < 0. \end{cases} \tag{49}$$

We tabulate $\mathscr{G}$ on a 400×400 regular grid spanning the range

$$x_0 = \log \min(p_0, p_{\max}) \tag{50}$$
$$x_{\min} = x_0 - (1 + \sqrt{2})\alpha \tag{51}$$
$$x_{\max} = \log p_{\max} \tag{52}$$
$$y_{\min} = 2x_0 - \sqrt{2}\alpha - x_{\max} \tag{53}$$
$$y_{\max} = x_0 + \alpha \tag{54}$$

where $\alpha = 4$ is a constant parameter that determines the extent of the grid.

2. *Lookup table construction*

The lookup table for $\mathscr{G}$ is populated as follows. If we neglect both the Bessel function and the r^m prior, then the approximate likelihood $\exp(-(p/r - p/r_0)^2)$ is maximized when $r = r_0$. The likelihood takes on a factor η (say, $\eta = 0.01$) of its maximum value when

$$r = r_\pm = \left(\frac{1}{r_0} \mp \frac{\sqrt{-\log \eta}}{p}\right)^{-1}. \tag{55}$$

We have now identified up to five breakpoints that partition the distance integrand into up to four intervals with quantitatively distinct behavior. These intervals are depicted in Figure 6 with distance increasing from left to right. There is a left-hand or small distance tail in which the integrand is small and monotonically increasing, a left- and right-hand side of the maximum likelihood peak, and a right-hand tail in which the integrand is small and monotonically decreasing. These breakpoints are:

$$r_{\text{break}} = \{r \in \begin{Bmatrix} r_{\min} \\ r_- \\ r_0 \\ r_+ \\ r_{\max} \end{Bmatrix} : r_{\min} \leq r \leq r_{\max}\}. \tag{56}$$

We use these breakpoints as initial subdivisions in an adaptive Gaussian quadrature algorithm[12]. This function estimates the integral over each subdivision and each interval's contribution to the total error, then subdivides the interval whose error contribution is largest. Subdivisions continue until a fixed total fractional error is reached. In this way, most integrand evaluations are expended on the most important distance interval, whether that happens to be the tails (when the posterior is dominated by the prior) or the peak (when the posterior is dominated by the observations).

3. *Interpolation*

The interpolant is evaluated slightly differently depending on which of the three regions marked I, II, and III in Figure 5 contains the point of interest. In region I, we use bicubic interpolation of $\log \mathcal{G}$ in x and y. In region II, we use univariate cubic interpolation of $\log \mathcal{G}$

[11] http://dlmf.nist.gov/10.40.E1

[12] for instance, GNU Scientific Library (GSL)'s `gsl_integrate_qagp` function, http://www.gnu.org/software/gsl/manual/html_node/QAGP-adaptive-integration-with-known-singular-points.html

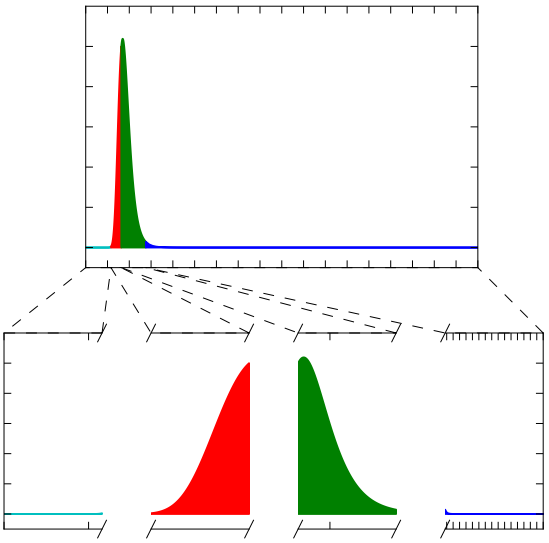

FIG. 6. Illustration of initial subdivisions for distance integration scheme. Distance increases from left to right. In the color version, the left-hand tail, the left- and right-hand sides of the maximum likelihood peak, and the right-hand tail, are colored cyan, red, green, and blue, respectively.

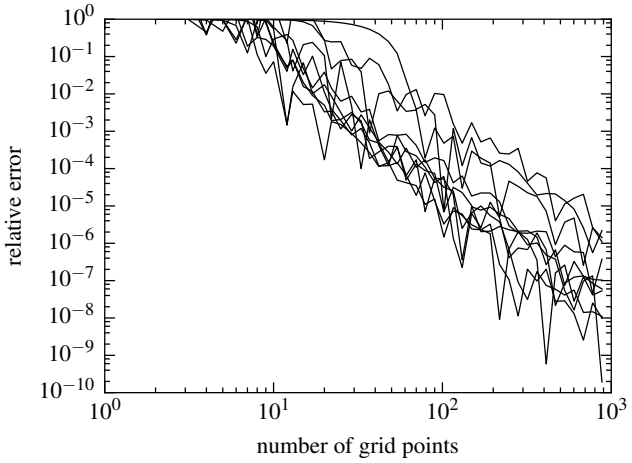

FIG. 7. Relative error in BAYESTAR distance integral interpolation scheme as a function of the size of the grid.

in x, with the sample points taken from the horizontal boundary between regions I and II. In region III, we use univariate cubic interpolation of $\log \mathcal{G}$ in $u = (x - y)/2$, with the sample points taken from the downward diagonal boundary between regions I and III. Finally, the distance integral $\mathcal{F}$ is obtained by multiplying the interpolated value of $\mathcal{G}$ by $\exp\left(p^2/{r_0}^2\right)$. For a 400×400 grid, the entire lookup table scheme is accurate to a relative error of about 10^{-5} in $\mathcal{F}$ (see Figure 7).

VI. ADAPTIVE HEALPIX SAMPLING

We have explained how we evaluate the marginal posterior at a given sky location. Now we must specify where we choose to evaluate it.

Our sampling of the sky relies completely on the Hierarchical Equal Area isoLatitude Pixelization (HEALPix, [64]), a special data structure designed for all-sky maps. HEALPix divides the sky into equal-area pixels. There is a hierarchy of HEALPix resolutions. A HEALPix resolution may be designated by its order N. The $N = 0$th order or base tiling has 12 pixels. At every successive order, each tile is subdivided into four new tiles. A resolution may also be referred to by the number of subdivisions along each side of the base tiles, $N_{\mathrm{side}} = 2^N$. There are $N_{\mathrm{pix}} = 12{N_{\mathrm{side}}}^2$ pixels at any given resolution. The HEALPix projection uniquely specifies the coordinates of the center of each pixel by providing a mapping from the resolution and pixel index $(N_{\mathrm{side}}, i_{\mathrm{pix}})$ to right ascension and declination (α, δ).

The BAYESTAR adaptive sampling process works as follows. We begin by evaluating the posterior probability density at the center of each of the $N_{\mathrm{pix},0} = 3072$ pixels of an $N_{\mathrm{side},0} = 16$ HEALPix grid. At this resolution, each pixel has an area of 13.4 deg^2. We then rank the pixels by contained probability (assuming constant probability density within a pixel), and subdivide the most probable $N_{\mathrm{pix},0}/4$ pixels into $N_{\mathrm{pix},0}$ new daughter pixels. We then evaluate the posterior again at the centers of the new daughter pixels, sort again, and repeat seven times. After the last iteration, we have evaluated the posterior probability density a total of $8N_{\mathrm{pix},0}$ times. On *most* subdivision steps, we descend one level deeper in HEALPix resolution. This process is illustrated in Figure 8.

The resulting map is a tree structure that describes a mesh of pixels with different resolutions. An example BAYESTAR subdivision is shown in Figure 9. To convert this mesh into a FITS image, we traverse the tree and flatten it into the highest resolution represented. The highest possible resolution is $N_{\mathrm{side}} = 2^{11}$, with an area of $\approx 10^{-3}$ deg^2 per pixel.[13]

VII. PARALLELIZATION

MCMC and similar stochastic schemes are typically very resistant to parallelization. However, BAYESTAR

[13] Although the resulting sky map contains $N_{\mathrm{pix}} \approx 5 \times 10^6$ pixels, at most $\approx 2 \times 10^4$ pixels have distinct values. For the purpose of delivery to observers, therefore, the output is always `gzip`-compressed with a ratio of $\approx 250 : 1$.

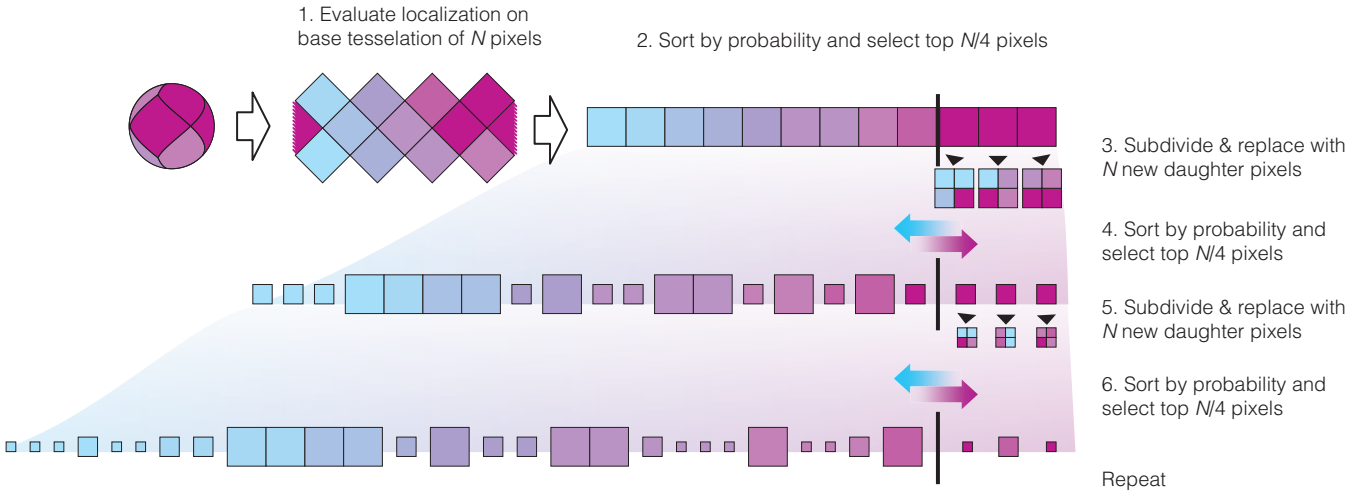

FIG. 8. Illustration of the BAYESTAR adaptive HEALPix sampling scheme.

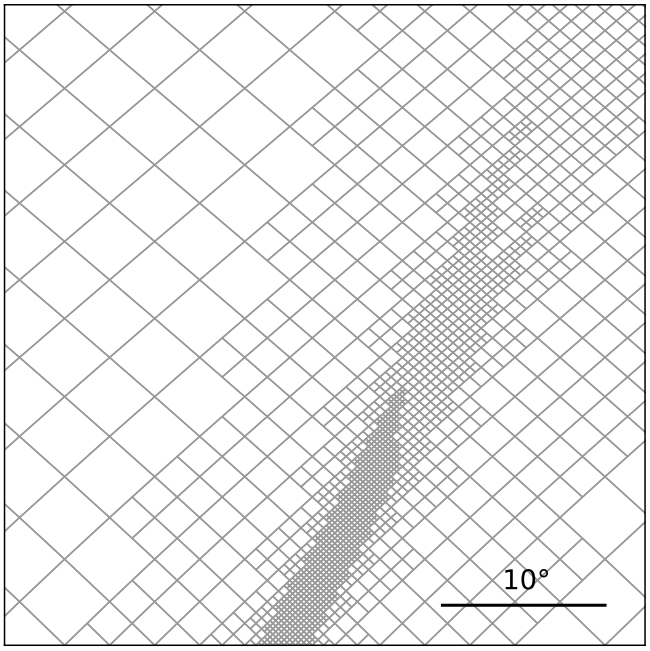

FIG. 9. An example multi-resultion HEALPix mesh arising from the BAYESTAR sampling scheme (plotted in a cylindrical projection). This is event 18951 from [27].

is completely deterministic and easily parallelizable because each pixel can be evaluated independently from all of the others. BAYESTAR consists of nine computationally intensive loops: the generation of the distance integral lookup table and the eight loops over pixels in the adaptive HEALPix sampling step. The iterations of each loop are distributed across multiple cores using OpenMP[14] compiler directives. In Section VIII E, we will show that BAYESTAR's run time is almost perfectly proportional to the number of cores, demonstrating that the serial sections (the sorts between the adaptation steps) are a negligible contribution to the overall wall clock time.

[14] http://openmp.org/

VIII. CASE STUDY

We have completed our description of the BAYESTAR algorithm. In [27], the authors presented a comprehensive and astrophysically realistic sample of simulated BNS mergers. We focused on the first two planned Advanced LIGO and Virgo observing runs as desribed in [3]. That work presented a catalog of 500 sky localizations from BAYESTAR and LALINFERENCE, and dealt with the quantitave position reconstruction accuracy as well as the qualitative sky morphologies. In the present work, we will use the same data set, but instead focus on demonstrating the correctness and performance of the BAYESTAR algorithm.

A. Observing scenarios

To review the assumptions made in [27], the two scenarios are:

2015 The first Advanced LIGO observing run, or "O1", scheduled to start in September 2015 and continue for three months. There are only two detectors participating in this run: LIGO Hanford ("H") and LIGO Livingston ("L"). Both detectors are expected to operate with a direction-averaged BNS merger range of 40–80 Mpc (though ongoing Advanced LIGO commissioning suggests that the higher end of this range will be achieved). As a result of having only two detectors, most localizations are long, thin arcs a few degrees wide and tens to hundreds of degrees long. The median 90%

credible area is about 600 deg^2.

2016 The second observing run, "O1", with the two between Advanced LIGO detectors, upgraded to a BNS range of 80–120 Mpc, operated jointly with the newly commissioned Advanced Virgo detector ("V"), operating at a range of 20–60 Mpc. The run is envisioned as lasting for six months in 2016–2017. The detectors are assumed to have random and independent 80% duty cycles. Consequently, all three detectors ("HLV") are in science mode about half of the time, with the remaining time divided roughly equally between each of the possible pairs ("HL", "HV", or "LV") and one or fewer detectors (at least two GW facilities are required for a detection). Virgo's range is assumed to be somewhat less than LIGO's because its commissioning timetable is about a year behind. Although the simulated signals are generally too weak in Virgo to trigger the matched-filter pipeline and contribute to *detection*, even these sub-threshold signals aid in position reconstruction with LALINFERENCE by lifting degeneracies. As a result, the median 90% credible area decreases to about 200 deg^2.

All simulated sources have component masses distributed uniformly between 1.2 and 1.6 $M_\odot$ and randomly oriented spins with dimensionless magnitudes $\chi = c|\mathbf{S}|/Gm^2$ between -0.05 and +0.05. Sky positions and binary orientations are random and isotropic. Distances are drawn uniformly from ${D_\mathrm{L}}^3$, reflecting a uniform source population (neglecting cosmological effects, which are small within the Advanced LIGO BNS range).

B. Detection and localization

The simulated waveforms were deposited in Gaussian noise that has been filtered to have the PSDs consistent with [3]. They were detected using the real-time matched-filter pipeline, GSTLAL_INSPIRAL [65]. Candidates with estimated false alarm rates (FARs) less than $10^{-2}\,\mathrm{yr}^{-1}$ were considered to be "detections." Because using Gaussian noise results in lower FARs than would be calculated in realistically glitchy detector noise, we imposed an additional detection threshold on the network S/N, $\hat{\rho} \geq 12$, which has been found to correspond to a comparable FAR in the initial LIGO runs.[15] Localizations for the "detections" were generated with BAYESTAR as well as the functionally equivalent and interchangeable LALINFERENCE_MCMC, LALINFERENCE_NEST, and LALINFERENCE_BAMBI samplers (collectively referred to as LALINFERENCE).

[15] See [61] for an analysis of the effect of glitchy noise on detection and parameter estimation.

C. Areas

We measured sky localization areas for each event as follows. First, we ranked the HEALPix pixels by descending posterior probability. Then, we computed the cumulative sum of the pixels in that order. Finally, we searched for the index of the pixel whose cumulative sum was equal to a given value: for example, 0.9 if we are interested in the 90% credible area. That pixel index times the area per pixel is the area of the smallest region of the specified credible level. This area can be thought of measuring the precision of the sky localization: it is a measure of the scale of the posterior distribution.

We can construct a second measure, called the searched area, as the smallest such constructed area that contains the true location of the source. A telescope with a FOV that is small compared to the characteristic scale of the posterior would intercept the true location of the source after covering the searched area. This measure is mainly useful because it measures the accuracy of the sky localization independently of the precision. In other words, it treats the sky map as merely a ranking statistic.

Histograms of the 90% credible area and the searched area are shown in Figure 10, broken down by observing scenario (2015 or 2016) and detector network (HL, HV, LV, or HLV). Note that there are no stastically significant differences in areas between BAYESTAR and LALINFERENCE, with the exception in the 2016/HLV configuration, for which some LALINFERENCE sky maps span about an order of magnitude less area than BAYESTAR. If we consider only events for which all three detectors contained a signal that was loud enough to trigger the matched-filter pipeline, the difference becomes much smaller and insignificant within 95% error bars. This is because if the signal is too weak to trigger the detection pipeline in one of the detectors, then BAYESTAR receives no information about that detector. This issue does not occur in the two-detector configurations (HL, HV, or LV) because two or more triggers are required to report a detection candidate.

This is a significant issue for the 2016 configuration, because the most accurate localizations are possible when all three detectors are operating. However, there may be a simple remedy. As we noted in Section III A, the BAYESTAR likelihood can be modified to use, instead of the times, phases, and amplitudes on arrival, the full complex matched-filter time series from all detectors. The detection pipeline, GSTLAL_INSPIRAL, would have to be modifed to save and transmit a small interval of the complex S/N time series (perhaps a few tens of milliseconds) around the time of each detection candidate. In addition to supplying the missing information for sub-threshold signals, this would make BAYESTAR mathematically equivalent to the LALINFERENCE analysis, but with the intrinsic parameters fixed to their maximum-likelihood values. This idea will be pursued in future work.

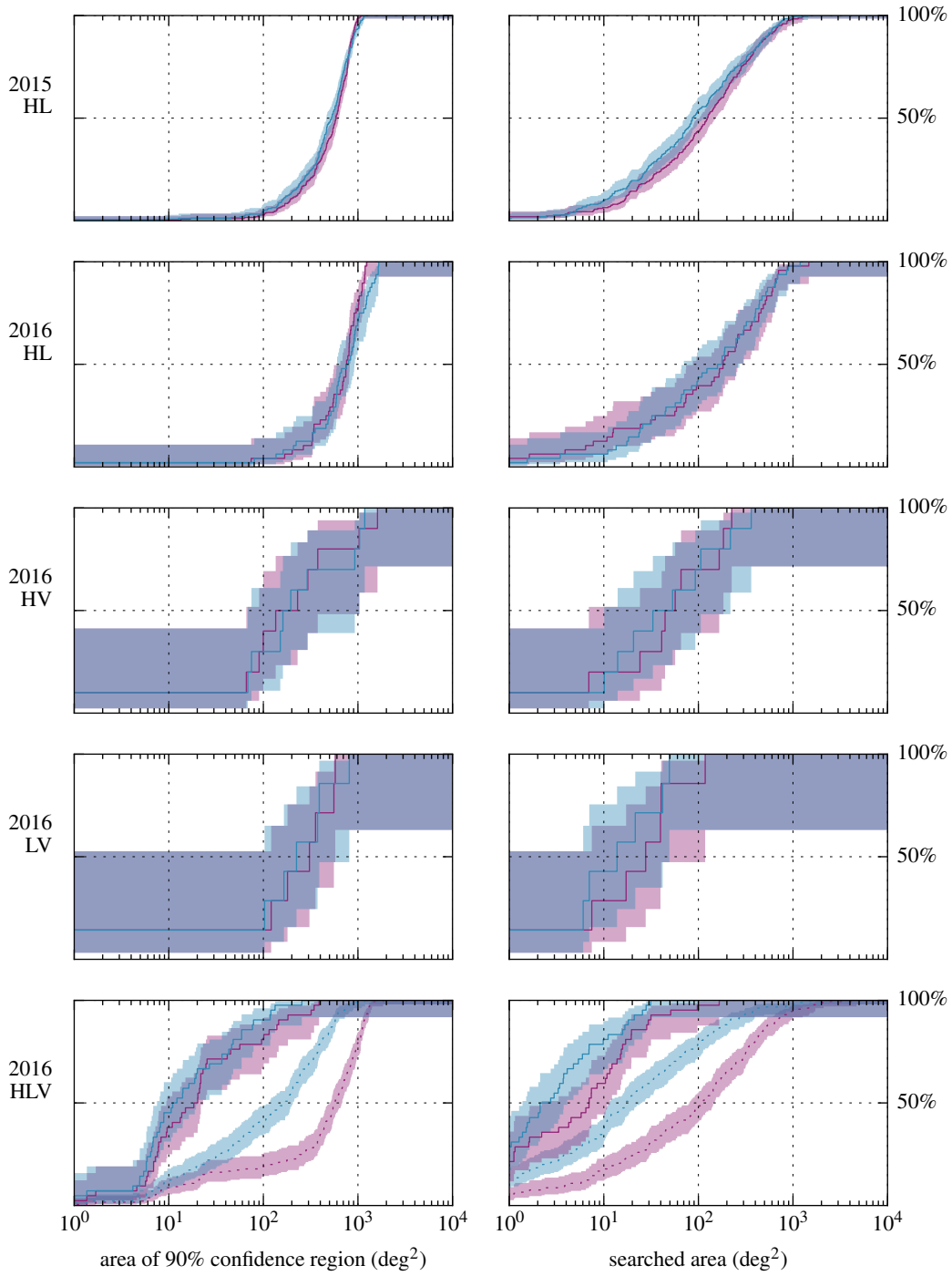

FIG. 10. Cumulative histograms of sky area, broken down by observing run and detector network. The plots in the left column show the 90% credible area and the plots in the right column show the searched area. From top bottom, the rows refer to the following observing scenarios/network configurations: 2015/HL, 2016/HL, 2016/HV, 2016/LV, 2016/HLV. The shaded regions represent the 95% confidence bounds. The magenta lines are represent BAYESTAR and the blue lines LALINFERENCE. Where relevant, dotted lines show all events in the given network configuratoin and solid lines show only events for which the matched-filter pipeline triggered on all operating detectors. Note that statistically significant differences in areas between the BAYESTAR and LALINFERENCE localizations occur only for events that were below the detection threshold in one or more detectors.

D. Self consistency

As we observed above, the area of a given credible region describes the precision of the sky localization, whereas the searched area describes the accuracy. However, self consistency requires that the two are related. For example, we should find that on average 90% of events have their true locations contained within their respective 90% credible regions. More generally, if we make a cumulative histogram of the credible levels corresponding to the searched areas of all of the events, then we should obtain a diagonal line (with small deviations due to finite sample size). This test, popularized for GW data analysis by [26], is a necessary but not sufficient condition for the validity of any Bayesian parameter estimation scheme.

It is already well established that LALINFERENCE localizations satisfy the P–P plot test when deployed with accurate templates and reasonable priors. We found at first that BAYESTAR's P–P plots tended to sag below the diagonal, indicating that though the accuracy (i.e., searched area) was comparable to LALINFERENCE, the *precision* was overstated, with confidence intervals that were only about 70% of the correct area. This was rectified by pre-scaling the S/Ns from GSTLAL_INSPIRAL by a factor of 0.83 prior to running BAYESTAR. This correction factor suggests that, for example, an S/N 10 trigger from GSTLAL_INSPIRAL has the *effective information content* of an S/N 8.3 signal. The missing information may be due to losses from the discreteness of the template bank, from the singular value decomposition (SVD), from mismatch between the matched-filter templates and the simulated signals, from the small but nonzero correlations between masses and intrinsic parameters, or from elsewhere within the detection pipeline. The correction is hard-coded into the rapid localization. With it, the P–P plots are diagonalized without negatively affecting the searched area (see Figure 11).

E. Run time

Since BAYESTAR is designed as one of the final steps in the real-time BNS search, it is important to characterize how long it takes to calculate a sky map. We compiled BAYESTAR with the Intel C Compiler (`icc`) at the highest architecture-specific optimization setting (`-ipo -O3 -xhost`). We timed it under Scientific Linux 6.1 on a Supermicro SuperServer 6028TP-HTTR system with dual 8-core Intel Xeon E5-2630 v3 CPUs clocked at 2.40 GHz, capable of executing 32 threads simultaneously (with hyperthreading). In Figure 12, we show how long it took to calculate a localization with BAYESTAR as the number of OpenMP threads was varied from 1 to 32. This is a violin plot, a smoothed vertical histogram. The magenta regions show run times for a two-detector network (HL) modeled on the first scheduled Advanced LIGO observing run in 2015, and the blue regions show run times for a three-detector network (HLV) based on the second planned observing run in 2016. These are the two observing scenarios that are discussed in [27].

Several features are apparent. First, at any number of threads, the two configurations have similar run times, although the 2016 events contain a subpopulation of outliers that take about 2.5 times as long as the 2015 events. These are probably due to taking one of the more expensive code branches in the distance integral interpolation. Second, the run times decrease proportionally to the number of threads. Based on experiences running BAYESTAR on the 32-core (64 threads with hyperthreading) cluster login machine, we expect the almost ideal parallel speedup to continue on machines with even more processors.

With just one thread, the BAYESTAR analysis takes 76–356 s, already orders of magnitude faster than the full parameter estimation. With 32 threads, BAYESTAR takes just 4–13 s. In practice, BAYESTAR's data handling (reading the detectors' PSDs, communicating with the GW candidate database, writing FITS files) take an additional 15 s or so, though this overhead could be reduced by parallelizing many of these steps. The overall latency is comparable to the other stages (data aggregation, trigger generation, alert distribution) in the real-time BNS analysis; therefore any significant further speedup would require significant changes through Advanced LIGO computing and infrastructure. The 32-thread configuration is representative of how BAYESTAR might be deployed in early Advanced LIGO.[16] For comparison, sky localization with LALINFERENCE takes about 100 hours [61].

Note that this benchmark shows BAYESTAR to be an order of magnitude faster than what was reported in [60, 61] due to the changes in the distance integration scheme that we noted in Section V B.

IX. FUTURE WORK

One immediately pressing direction for future work is to address the issue of sub-threshold signals, as this will be a major issue when Advanced Virgo comes online in 2016–2017. Using the full S/N time series in place of the autocorrelation function seems like a promising avenue; implementing this requires some infrastructure changes to both the matched filter pipeline and BAYESTAR. Along these lines, we also refer the reader to [66] for a similar, non-MCMC approach to rapid exploration of the full parameter space.

[16] BAYESTAR has been successfully ported to the Intel's Many Integrated Core (MIC) architecture and has been tested in a 500 thread configuration on a system with dual Intel Xeon Phi 5110P coprocessors.

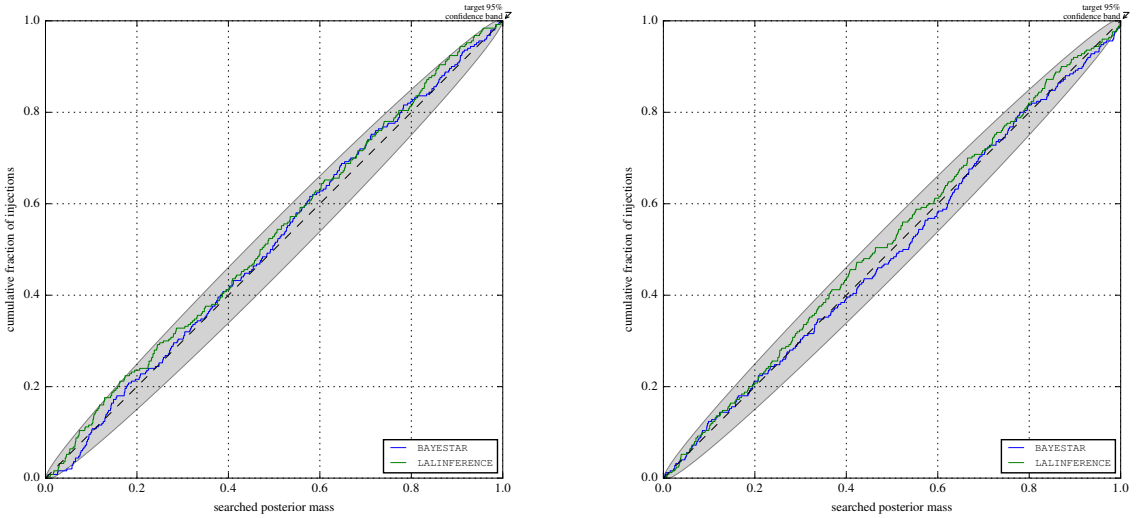

FIG. 11. P–P plots for BAYESTAR and LALINFERENCE localizations in the 2015 and 2016 configurations. The gray lozenge around the diagonal is a target 95% confidence band derived from a binomial distribution.

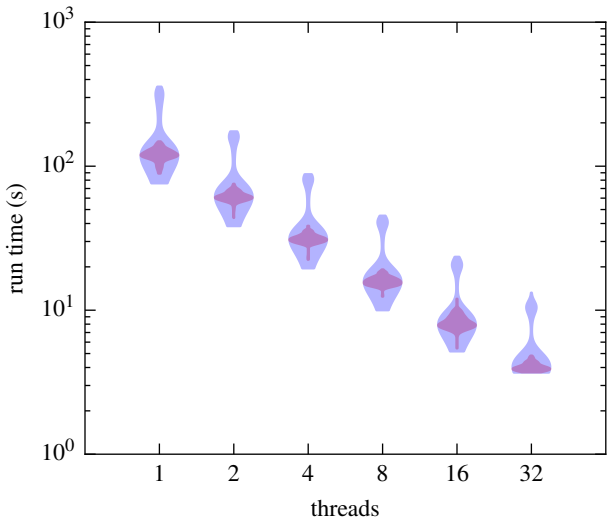

FIG. 12. Violin plot of BAYESTAR run times as the number of OpenMP threads is varied from 1 to 32. The 2015 scenario is shown in red and the 2016 scenario in blue.

A more open-ended question is how to account for spin precession. The simulations in [27] and in this paper featured extremely modest spins of $\chi \leq 0.05$, consistent with the fastest known pulsars in binaries [67, 68]. The signals were detected using a template bank that lacked spins entirely. Farr et al. (in preparation) show that using nonspinning BNS templates for parameter estimation has negligible impact on sky localization. However, if one or both companions are spinning as fast as a millisecond pulsar, $\chi \sim 0.4$ [69], or even near breakup, $\chi \sim 0.7$, then the orbital plane may precess; in this case, spins can no longer be neglected for detection [68] and may also be important for parameter estimation. Since spinning BNS searches are still an active area of development, BAYESTAR's sky localization accuracy in this regime should be reexamined in the future.

Although the response time of BAYESTAR has been driven by the anticipated timescales for kilonova and afterglow emission, a recurring question is whether there is any detectable EM signal in the seconds before, during, and after the merger itself. Since the GW inspiral signal is in principal detectable for up to hundreds of seconds before merger, one could imagine positioning rapidly slewing instruments to search for any prompt emission. This concept was explored by the authors [65], and will be revisted in further detail in both Sathyaprakash et al. (in preparation) and Chu et al. (in preparation).

On the topic of very low-latency localization, we also recommend Chen & Holz (in preparation), who propose a rapid localization scheme that is similar to ours, but even faster because it makes some additional compromises: their likelihood is strictly Gaussian, so one more marginalization integral (the integral over arrival time) can be performed analytically.

X. CONCLUSION

We have presented a novel, fast, accurate, Bayesian algorithm for inferring the sky locations of compact bi-

nary merger sources that may soon be detected by advanced ground-based GW detectors. For BNS systems with small spins, we have shown that BAYESTAR produces sky maps that are as accurate as the full MCMC parameter estimation code, but can do so within ∼10 seconds after a detection. Still faster response times should be possible in the future (if warranted) by deploying BAYESTAR on machines with more cores, or by distributing BAYESTAR across multiple computers.

Following a BNS merger, the signal will be detected by the matched-filter pipeline within tens of seconds; an alert containing the time and estimated significance of the event can be distributed almost immediately (although a human validation stage that may be present at the beginning of the first observing run may introduce some additional latency). The localization from BAYESTAR will be available tens of seconds to a minute later. Finally, the refined localiation and the detailed estimates of masses and spins from LALINFERENCE will be distributed hours to days later.

Relevant timescales for possible EM counterparts to GW signals include seconds (the prompt GRB signature), hundreds of seconds (extended emission and X-ray plateaus that are observed for some short GRBs), minutes to hours (X-ray and optical afterglow), hours to days (the kilonova or the blue flashes associated with unbound ejecta or disk winds), and days to years (the radio afterglow). For the first time, we are able to provide accurate localizations before the peak of any of these EM signatures (except for the short GRB or any pre-merger signal). Even for components like the kilonova that should peak within hours to days, the availability of the localizations within seconds might provide a window of several hours to obtain tiled images of the area before the EM emission begins. These could be used as reference images, crucial at optical wavelengths for establishing the rapid rise and quickly distinguishing from slower background transients.

ACKNOWLEDGMENTS

This paper has LIGO Document Number LIGO-P1500009-v5.

We thank John Veitch and Will Farr for chairing a review of the analysis and code. We thank Britt Griswold for assistance with preparing Figure 8.

LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the National Science Foundation (NSF) and operates under cooperative agreement PHY-0107417.

This research was supported by the NSF through a Graduate Research Fellowship to L.S. L.S. thanks the Aspen Center for Physics and NSF Grant #1066293 for hospitality during the editing of this paper.

Source code for BAYESTAR is available as part of LALInference[17], the open source LIGO/Virgo parameter estimation toolchain, which is in turn part of LALSuite[18]. This research made use of Astropy[19] [70], a community-developed core Python package for astronomy. Some of the results in this paper have been derived using HEALPix[20] [64].

Some results were produced on the NEMO computing cluster operated by the Center for Gravitation and Cosmology at University of Wisconsin–Milwaukee under NSF Grants PHY-0923409 and PHY-0600953.

Appendix A: Independence of intrinsic and extrinsic errors

If all of the detectors have the same noise PSDs up to multiplicative factors, $c_1 S_1(\omega) = c_2 S_2(\omega) = \cdots = c_n S_n(\omega) \equiv S(\omega)$, then we can show that the errors in the intrinsic parameters (masses) are not correlated with sky position errors. This is because we can change variables from amplitudes, phases, and times to amplitude ratios, phase differences, and time differences. With N detectors, we can form a single average amplitude, time, and phase, plus $N-1$ linearly independent differences. The averages are correlated with the intrinsic parameters, but neither are correlated with the differences. Since only the differences inform sky location, this gives us license to neglect uncertainty in masses when we are computing sky resolution.

This is easiest to see if we make the temporary change of variables $\rho \to \varsigma = \log \rho$. This allows us to factor out the S/N dependence from the single-detector Fisher matrix. The extrinsic part becomes

$$\begin{aligned}\mathcal{I}_{\boldsymbol{\theta}_i,\boldsymbol{\theta}_i} &= \begin{array}{c} \\ \varsigma_i \\ \gamma_i \\ \tau_i \end{array} \begin{array}{c} \begin{array}{ccc} \varsigma_i & \gamma_i & \tau_i \end{array} \\ \begin{pmatrix} {\rho_i}^2 & 0 & 0 \\ 0 & {\rho_i}^2 & -{\rho_i}^2\overline{\omega}_i \\ 0 & -{\rho_i}^2\overline{\omega}_i & {\rho_i}^2\overline{\omega^2}_i \end{pmatrix} \end{array} \\ &= {\rho_i}^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -\overline{\omega}_i \\ 0 & -\overline{\omega}_i & \overline{\omega^2}_i \end{pmatrix}. \end{aligned} \tag{A1}$$

Due to our assumption that the detectors' PSDs are proportional to each other, the noise moments are the same for all detectors, $\overline{\omega^k}_i \equiv \overline{\omega^k}$. Then we can write the single-detector Fisher matrix as

$$\mathcal{I}_i = {\rho_i}^2 \begin{pmatrix} A & B \\ B^{\mathsf{T}} & C \end{pmatrix}, \tag{A2}$$

[17] https://ligo-vcs.phys.uwm.edu/cgit/lalsuite/tree/lalinference
[18] http://www.lsc-group.phys.uwm.edu/daswg/projects/lalsuite.html
[19] http://www.astropy.org
[20] http://healpix.sourceforge.net

with the top-left block A comprising the extrinsic parameters and the bottom-right block C the intrinsic parameters.

Information is additive, so the Fisher matrix for the whole detector network is

$$\mathcal{I}_{\rm net} = \begin{pmatrix} {\rho_1}^2 A & 0 & \cdots & 0 & {\rho_1}^2 B \\ 0 & {\rho_2}^2 A & & \vdots & {\rho_1}^2 B \\ \vdots & & \ddots & 0 & \vdots \\ 0 & 0 & \cdots & {\rho_N}^2 A & {\rho_N}^2 B \\ {\rho_1}^2 B^{\mathsf{T}} & {\rho_2}^2 B^{\mathsf{T}} & \cdots & {\rho_N}^2 B^{\mathsf{T}} & {\rho_{\rm net}}^2 C \end{pmatrix}. \quad \text{(A3)}$$

Now we introduce the change of variables that sacrifices the Nth detector's extrinsic parameters for the network averages,

$$\begin{array}{rcl} \varsigma_N & \to & \overline{\varsigma} = \left(\sum_i {\rho_i}^2 \varsigma_i\right) / {\rho_{\rm net}}^2, \\ \gamma_N & \to & \overline{\gamma} = \left(\sum_i {\rho_i}^2 \gamma_i\right) / {\rho_{\rm net}}^2, \\ \tau_N & \to & \overline{\tau} = \left(\sum_i {\rho_i}^2 \tau_i\right) / {\rho_{\rm net}}^2, \end{array} \quad \text{(A4)}$$

and replaces the first $N-1$ detectors' extrinsic parameters with differences,

$$\left.\begin{array}{rcl} \varsigma_i & \to & \delta\varsigma_i = \varsigma_i - \overline{\varsigma} \\ \gamma_i & \to & \delta\gamma_i = \gamma_i - \overline{\gamma} \\ \tau_i & \to & \delta\tau_i = \tau_i - \overline{\tau} \end{array}\right\} \text{ for } i = 1, \ldots, N-1. \quad \text{(A5)}$$

The Jacobian matrix that describes this change of variables is

$$J = \begin{pmatrix} 1 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 1 & & 0 & 1 & 0 \\ \vdots & & \ddots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 1 & 0 \\ \frac{-{\rho_1}^2}{{\rho_N}^2} & \frac{-{\rho_2}^2}{{\rho_N}^2} & \cdots & \frac{-{\rho_{N-1}}^2}{{\rho_N}^2} & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}. \quad \text{(A6)}$$

The transformed network Fisher matrix is block diagonal,

$$\mathcal{I}_{\rm net} \to J^{\mathsf{T}} \mathcal{I}_{\rm net} J = \begin{pmatrix} {\rho_1}^2\left(1+\frac{1}{{\rho_1}^4}\right)A & \frac{{\rho_1}^2{\rho_2}^2}{{\rho_N}^2}A & \cdots & \frac{{\rho_1}^2{\rho_{N-1}}^2}{{\rho_N}^2}A & 0 & 0 \\ \frac{{\rho_1}^2{\rho_2}^2}{{\rho_N}^2}A & {\rho_2}^2\left(1+\frac{1}{{\rho_1}^4}\right)A & & \frac{{\rho_2}^2{\rho_{N-1}}^2}{{\rho_N}^2}A & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots & \vdots \\ \frac{{\rho_1}^2{\rho_{N-1}}^2}{{\rho_N}^2}A & \frac{{\rho_2}^2{\rho_{N-1}}^2}{{\rho_N}^2}A & \cdots & {\rho_{N-1}}^2\left(1+\frac{1}{{\rho_1}^4}\right)A & 0 & 0 \\ 0 & 0 & \cdots & 0 & {\rho_{\rm net}}^2 A & {\rho_{\rm net}}^2 B \\ 0 & 0 & \cdots & 0 & {\rho_{\rm net}}^2 B^{\mathsf{T}} & {\rho_{\rm net}}^2 C \end{pmatrix}. \quad \text{(A7)}$$

The top-left block contains $N-1$ relative amplitudes, phases, and times on arrival, all potentially correlated with each other. The bottom-right block contains the average amplitudes, phases, and times, as well as the masses. The averages and the masses are correlated with each other, but are not correlated with the differences. Because only the differences are informative for sky localization, we drop the intrinsic parameters from the rest of the Fisher matrix calculations in this appendix.

Appendix B: Interpretation of phase and time errors

The Fisher matrix in Equation (23) is block diagonal, which implies that estimation errors in the signal amplitude ρ are uncorrelated with the phase γ and time τ. A sequence of two changes of variables lends some physical interpretation to the nature of the coupled estimation errors in γ and τ.

First, we put the phase and time on the same footing by measuring the time in units of $1/\sqrt{\overline{\omega^2}}$ with a change of variables from τ to $\gamma_\tau = \sqrt{\overline{\omega^2}}\tau$:

$$\mathcal{I}' = \begin{array}{c} \\ \rho_i \\ \gamma_i \\ \gamma_{\tau,i} \end{array} \begin{array}{c} \begin{array}{ccc} \rho_i & \gamma_i & \gamma_{\tau,i} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & {\rho_i}^2 & -{\rho_i}^2 \frac{\overline{\omega}_i}{\sqrt{\overline{\omega^2}_i}} \\ 0 & -{\rho_i}^2 \frac{\overline{\omega}_i}{\sqrt{\overline{\omega^2}_i}} & {\rho_i}^2 \end{pmatrix} \end{array}. \quad \text{(B1)}$$

The second change of variables, from γ and γ_τ to $\gamma_\pm = \frac{1}{\sqrt{2}}(\gamma \pm \gamma_\tau)$, diagonalizes the Fisher matrix:

$$\mathcal{I}'' = \begin{array}{c} \\ \rho_i \\ \gamma_{+,i} \\ \gamma_{-,i} \end{array} \begin{array}{c} \begin{array}{ccc} \rho_i & \gamma_{+,i} & \gamma_{-,i} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & \left(1 - \frac{\overline{\omega}_i}{\sqrt{\overline{\omega^2}_i}}\right){\rho_i}^2 & 0 \\ 0 & 0 & \left(1 + \frac{\overline{\omega}_i}{\sqrt{\overline{\omega^2}_i}}\right){\rho_i}^2 \end{pmatrix} \end{array}. \quad \text{(B2)}$$

Thus, in the appopriate time units, the *sum and difference* of the phase and time of the signal are measured independently.

Appendix C: Position resolution

Finally, we will calculate the position resolution of a network of GW detectors. We could launch directly into computing derivatives of the full signal model from Equation (6) with respect to all of the parameters, but this would result in a very complicated expression. Fortunately, we can take two shortcuts. First, since we showed in Section A that the intrinsic parameters are correlated only with an overall nuisance average arrival time, amplitude, and phase, we need not consider the derivatives with respect to mass at all. Second, we can reuse the extrinsic part of the single detector Fisher matrix from Equation (23) by computing the much simpler Jacobian matrix to transform from the time, amplitude, and phase on arrival, to the parameters of interest.

We begin by transforming the single-detector Fisher matrix from a polar to a rectangular representation of the complex amplitude given in Equations (14, 13), $\rho_i, \gamma_i \to \Re[z_i] = \rho_i \cos\gamma_i, \Im[z_i] = \rho_i \sin\gamma_i$:

$$\mathcal{I}_i = \begin{array}{c} \\ \Re[z_i] \\ \Im[z_i] \\ \tau_i \end{array} \begin{array}{c} \begin{array}{ccc} \Re[z_i] & \Im[z_i] & \tau_i \end{array} \\ \begin{pmatrix} 1 & 0 & \overline{\omega}_i b_i \\ 0 & 1 & -\overline{\omega}_i b_i \\ \overline{\omega}_i b_i & -\overline{\omega}_i b_i & {\rho_i}^2 \overline{\omega^2}_i \end{pmatrix} \end{array}. \tag{C1}$$

Consider a source in a "standard" orientation with the direction of propagation along the $+z$ axis, such that the GW polarization tensor may be written in Cartesian coordinates as

$$H = \frac{1}{r} e^{2i\phi_c} \begin{pmatrix} \frac{1}{2}(1+\cos^2\iota) & i\cos\iota & 0 \\ i\cos\iota & -\frac{1}{2}(1+\cos^2\iota) & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{C2}$$

Now introduce a rotation matrix R that actively transforms this source to the Earth-relative polar coordinates θ, ϕ, and gives the source a polarization angle ψ (adopting temporarily the notation $\mathrm{c}_\theta = \cos\theta$, $\mathrm{s}_\theta = \sin\theta$):

$$R = R_z(\phi) R_y(\theta) R_z(\psi) R_y(\pi) = \begin{pmatrix} \mathrm{c}_\phi & -\mathrm{s}_\phi & 0 \\ \mathrm{s}_\phi & -\mathrm{c}_\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathrm{c}_\theta & 0 & \mathrm{s}_\theta \\ 0 & 1 & 0 \\ -\mathrm{s}_\theta & 0 & \mathrm{c}_\theta \end{pmatrix} \begin{pmatrix} \mathrm{c}_\psi & -\mathrm{s}_\psi & 0 \\ \mathrm{s}_\psi & -\mathrm{c}_\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{C3}$$

(The rightmost rotation reverses the propagation direction so that the wave is traveling *from* the sky position θ, ϕ.) With the (symmetric) detector response tensor D_i, we can write the received amplitude and arrival time as

$$z_i = r_{1,i} \operatorname{Tr}\left[D_i R\, H\, R^{\mathsf{T}}\right], \tag{C4}$$

$$\tau_i = t_\oplus + \mathbf{d}_i^{\mathsf{T}} R\, \mathbf{k}. \tag{C5}$$

Equivalently, we can absorb the rotation R and the horizon distance $r_{1,i}$ into the polarization tensor, detector response tensors, and positions,

$$H \to H' = R_z(\psi)\, R_y(\pi)\, H\, R_y(\pi)^{\mathsf{T}}\, R_z(\psi)^{\mathsf{T}}, \tag{C6}$$

$$D_i \to D_i' = r_{1,i}\, R_y(\theta)^{\mathsf{T}}\, R_z(\phi)^{\mathsf{T}}\, D_i\, R_z(\phi)\, R_y(\theta), \tag{C7}$$

$$\mathbf{d}_i \to \mathbf{d}_i' = R_y(\theta)^{\mathsf{T}}\, R_z(\phi)^{\mathsf{T}}\, \mathbf{d}_i, \tag{C8}$$

$$\mathbf{k} \to \mathbf{k}' = (0,0,-1). \tag{C9}$$

Now the model becomes

$$H' = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{C10}$$

$$z_i = \operatorname{Tr}\left[D_i' H'\right] = h_+(D_{00}' - D_{11}') + 2h_\times D_{01}', \tag{C11}$$

$$\tau_i = t_\oplus + (\mathbf{d}_i') \cdot \mathbf{k}, \tag{C12}$$

where

$$h_+ = \frac{1}{r} e^{2i\phi_c} \left[\frac{1}{2}(1+\cos^2\iota)\cos 2\psi + i\cos\iota \sin 2\psi\right], \tag{C13}$$

$$h_\times = \frac{1}{r} e^{2i\phi_c} \left[\frac{1}{2}(1+\cos^2\iota)\sin 2\psi - i\cos\iota \cos 2\psi\right]. \tag{C14}$$

We insert an infinitesimal rotation δR to perturb the source's orientation from the true value:

$$z_i = \operatorname{Tr}\left[D_i'(\delta R) H'(\delta R)^{\mathsf{T}}\right], \tag{C15}$$

$$\tau_i = t_\oplus + (\mathbf{d}_i')^{\mathsf{T}}(\delta R)\mathbf{k}'. \tag{C16}$$

We only need a first order expression for δR, because we will be taking products of first derivatives of it[21]:

$$\delta R = \begin{pmatrix} 1 & 0 & \delta\theta \\ 0 & 1 & \delta\phi \\ -\delta\theta & -\delta\phi & 1 \end{pmatrix}. \tag{C17}$$

[21] Caution: the angles $\delta\theta$ and $\delta\phi$ represent displacements in two orthogonal directions, but are *not* necessarily simply related to θ and ϕ.

We construct a Jacobian matrix J_i to transform from the single-detector observables $(\Re[z_i], \Im[z_i], \tau_i)$ to the position perturbations, polarization components, and geocentered arrival time $(\delta\theta, \delta\phi, \Re[h_+], \Im[h_+], \Re[h_\times], \Im[h_\times], t_\oplus)$:

$$J_i^{\mathsf{T}} = \begin{array}{c} \\ \delta\theta \\ \delta\phi \\ \Re[h_+] \\ \Im[h_+] \\ \Re[h_\times] \\ \Im[h_\times] \\ t_\oplus \end{array} \begin{array}{c} \begin{array}{ccc} \Re[z_i] & \Im[z_i] & \tau_i \end{array} \\ \begin{pmatrix} -2\Re[h_+]D'_{02} - 2\Re[h_\times]D'_{12} & -2\Im[h_+]D'_{02} - 2\Im[h_\times]D'_{12} & -d'_0 \\ -2\Re[h_\times]D'_{02} + 2\Re[h_+]D'_{12} & -2\Im[h_\times]D'_{02} + 2\Im[h_+]D'_{12} & -d'_1 \\ D'_{00} - D'_{11} & 0 & 0 \\ 0 & D'_{00} - D'_{11} & 0 \\ 2D'_{01} & 0 & 0 \\ 0 & 2D'_{01} & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array}. \tag{C18}$$

We transform and sum the information from each detector,

$$\mathcal{I}_{\rm net} = \sum_i J_i^{\mathsf{T}} \mathcal{I}_i J_i. \tag{C19}$$

1. Marginalization over nuisance parameters

To extract an area from the Fisher matrix, we must first marginalize or discard the nuisance parameters. Note that marginalizing parameters of a multivariate Gaussian distribution amounts to simply dropping the relevant entries in the mean vector and covariance matrix. Since the information is the inverse of the covariance matrix, we need to invert the Fisher matrix, drop all but the first two rows and columns, and then invert again.

This procedure has a shortcut called the Schur complement (see, for example, Press *et al.* 71). Consider a partitioned square matrix M and its inverse:

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \qquad M^{-1} = \begin{pmatrix} \tilde{A} & \tilde{B} \\ \tilde{C} & \tilde{D} \end{pmatrix}. \tag{C20}$$

If A and B are square matrices, then the upper-left block of the inverse can be written as

$$\tilde{A}^{-1} = A - BD^{-1}C. \tag{C21}$$

If we partition the $\mathcal{I}_{\rm net}$ similarly, the A block consists of the first two rows and columns and D is the lower right block that describes all other parameters. Because the Fisher matrix is symmetric, the off-diagonal blocks satisfy $C = B^{\mathsf{T}}$. Then the Schur complement

$$\mathcal{I}_{\rm marg} = A - BD^{-1}B^{\mathsf{T}} \tag{C22}$$

gives us the information matrix marginalized over all parameters but $\delta\theta$ and $\delta\phi$.

2. Spatial interpretation

How do we extract the dimensions of the localization from the Fisher matrix? If there are $N \leq 2$ detectors, then the Fisher matrix must be degenerate, because there are $3N$ measurements and 7 parameters:

$$\left\{ \begin{array}{c} \delta\theta \\ \delta\phi \\ \Re[h_+] \\ \Im[h_+] \\ \Re[h_\times] \\ \Im[h_\times] \\ t_\oplus \end{array} \right\} = 7 \text{ parameters} \qquad \longleftrightarrow \qquad \left\{ \begin{array}{c} \Re[z_i] \\ \Im[z_i] \\ \tau_i \end{array} \right\} \times N = 3N \text{ observables.}$$

Therefore, for $N = 2$ detectors, the marginalized Fisher matrix $\mathcal{I}_{\rm marg}$ is singular. Its only nonzero eigenvalue λ describes the width of an annulus on the sky. The width of the annulus that contains probability p is given by

$$L_p = 2\sqrt{2}\,\mathrm{erf}^{-1}(p)/\sqrt{\lambda}. \tag{C23}$$

The prefactor $2\sqrt{2}/\,\mathrm{erf}^{-1}(p)$ is the central interval of a normal distribution that contains a probability p, and is ≈ 3.3 for $p = 0.9$. *Caution:* for two-detector networks, priors play an important role in practical parameter estimation and areas can be much smaller than one would

predict from the Fisher matrix.

For $N \geq 3$ detectors, the parameters are over-constrained by the data and the Fisher matrix describes the dimensions of an ellipse. Within a circle of radius r centered on the origin, the enclosed probability p is

$$p = \int_0^{2\pi} \int_0^r \frac{1}{2\pi} e^{-s^2/2} s \, ds \, d\phi = 1 - e^{-r^2/2}. \quad \text{(C24)}$$

Therefore the radius r of the circle that contains a probability p is

$$r = \sqrt{-2\ln(1-p)}. \quad \text{(C25)}$$

Suppose that the eigenvalues of the Fisher matrix are λ_1 and λ_2. This describes a 1σ uncertainty ellipse that has major and minor radii ${\lambda_1}^{-1/2}$, ${\lambda_2}^{-1/2}$, and area $A_{1\sigma} = \pi/\sqrt{\lambda_1\lambda_2} = \pi/\sqrt{\det \mathcal{I}}$. Then the area of an ellipse containing probability p is

$$A_p = -2\pi \ln(1-p)/\sqrt{\det \mathcal{I}}, \quad \text{(C26)}$$

or, more memorably for the 90th percentile, $A_{0.9} = 2\pi \ln(10)/\sqrt{\det \mathcal{I}}$.

3. Outline of calculation

Using the above derivation, we arrive at a prediction for the sky resolution of a GW detector network. We took some shortcuts that allowed us to avoid directly evaluating the complicated derivatives of the signal itself with respect to sky location. As a result, the expressions involved in each step are simple enough to be manually entered into a computer program. However, because the procedure involves several steps, we outline it once again below.

1. Compute, for each detector, the horizon distance $r_{1,i}$, the angular frequency moments $\overline{\omega}_i$ and $\overline{\omega^2}_i$, and $(h_+, h_\times)$ from Equations (C13, C14). (These can be reused for multiple source positions as long as the masses and the detector noise PSDs are the same.)

2. For a given ϕ, θ, ψ, compute the complex received amplitude z_i from Equations (C10, C11), the extrinsic Fisher matrix from Equation (23), and the Jacobian from Equation (C18).

3. Sum the information from all detectors using Equation C19.

4. Compute the marginalized Fisher matrix from the Schur complement using Equation (C22).

5. If there are two detectors, find the width L_p of the ring describing the pth quantile using Equation (C23). If there are three or more detectors, find the area A_p of the pth quantile using Equation (C26).
6. (Optionally, convert from (ste)radians to (square) degrees.)

See code listing in Appendix A.6 of [60].

[1] G. M. Harry and LIGO Scientific Collaboration, Classical and Quantum Gravity **27**, 084006 (2010).
[2] The LIGO Scientific Collaboration, J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, K. Ackley, C. Adams, T. Adams, P. Addesso, and et al., Classical and Quantum Gravity **32**, 074001 (2015), arXiv:1411.4547 [gr-qc].
[3] LIGO Scientific Collaboration, Virgo Collaboration, J. Aasi, J. Abadie, B. P. Abbott, R. Abbott, T. D. Abbott, M. Abernathy, T. Accadia, F. Acernese, and et al., ArXiv e-prints (2013), arXiv:1304.0670 [gr-qc].
[4] F. Acernese, M. Agathos, K. Agatsuma, D. Aisa, N. Allemandou, A. Allocca, J. Amarni, P. Astone, G. Balestri, G. Ballardin, and et al., Classical and Quantum Gravity **32**, 024001 (2015), arXiv:1408.3978 [gr-qc].
[5] K. Kuroda, International Journal of Modern Physics D **20**, 1755 (2011).
[6] Y. Aso, Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, and H. Yamamoto, Phys. Rev. D **88**, 043007 (2013), arXiv:1306.6747 [gr-qc].
[7] C. S. Unnikrishnan, International Journal of Modern Physics D **22**, 1341010 (2013).
[8] J. Abadie *et al.*, Class. Quantum Grav. **27**, 173001 (2010).
[9] B. D. Metzger and E. Berger, ApJ **746**, 48 (2012), arXiv:1108.6056 [astro-ph.HE].
[10] J. Barnes and D. Kasen, ApJ **775**, 18 (2013), arXiv:1303.5787 [astro-ph.HE].
[11] B. D. Metzger, A. Bauswein, S. Goriely, and D. Kasen, MNRAS **446**, 1115 (2015), arXiv:1409.0544 [astro-ph.HE].
[12] B. D. Metzger and R. Fernández, MNRAS **441**, 3444 (2014), arXiv:1402.4803 [astro-ph.HE].
[13] B. Paczynski, ApJ **308**, L43 (1986).
[14] D. Eichler, M. Livio, T. Piran, and D. N. Schramm, Nature **340**, 126 (1989).
[15] R. Narayan, B. Paczynski, and T. Piran, ApJ **395**, L83 (1992), astro-ph/9204001.
[16] L. Rezzolla, B. Giacomazzo, L. Baiotti, J. Granot, C. Kouveliotou, and M. A. Aloy, ApJ **732**, L6 (2011), arXiv:1101.4298 [astro-ph.HE].
[17] S. Fairhurst, New Journal of Physics **11**, 123006 (2009), arXiv:0908.2356 [gr-qc].
[18] L. Wen and Y. Chen, Phys. Rev. D **81**, 082001 (2010), arXiv:1003.2504.

[19] S. Fairhurst, Class. Quantum Grav. **28**, 105021 (2011), arXiv:1010.6192 [gr-qc].
[20] S. Vitale and M. Zanolin, Phys. Rev. D **84**, 104020 (2011), arXiv:1108.2410 [gr-qc].
[21] C. L. Rodriguez, B. Farr, V. Raymond, W. M. Farr, T. B. Littenberg, D. Fazi, and V. Kalogera, ApJ **784**, 119 (2014), arXiv:1309.3273 [astro-ph.HE].
[22] S. Nissanke, J. Sievers, N. Dalal, and D. Holz, ApJ **739**, 99 (2011), arXiv:1105.3184 [astro-ph.CO].
[23] S. Nissanke, M. Kasliwal, and A. Georgieva, ApJ **767**, 124 (2013).
[24] M. M. Kasliwal and S. Nissanke, ApJ **789**, L5 (2014), arXiv:1309.1554 [astro-ph.HE].
[25] K. Grover, S. Fairhurst, B. F. Farr, I. Mandel, C. Rodriguez, T. Sidery, and A. Vecchio, Phys. Rev. D **89**, 042004 (2014), arXiv:1310.7454 [gr-qc].
[26] T. Sidery, B. Aylott, N. Christensen, B. Farr, W. Farr, F. Feroz, J. Gair, K. Grover, P. Graff, C. Hanna, V. Kalogera, I. Mandel, R. O'Shaughnessy, M. Pitkin, L. Price, V. Raymond, C. Röver, L. Singer, M. van der Sluys, R. J. E. Smith, A. Vecchio, J. Veitch, and S. Vitale, Phys. Rev. D **89**, 084060 (2014), arXiv:1312.6013 [astro-ph.IM].
[27] L. P. Singer, L. R. Price, B. Farr, A. L. Urban, C. Pankow, S. Vitale, J. Veitch, W. M. Farr, C. Hanna, K. Cannon, T. Downes, P. Graff, C.-J. Haster, I. Mandel, T. Sidery, and A. Vecchio, ApJ **795**, 105 (2014), arXiv:1404.5623 [astro-ph.HE].
[28] S. R. Kulkarni, arXiv:1202.2381 (2012).
[29] J. Abadie, B. P. Abbott, R. Abbott, T. D. Abbott, M. Abernathy, T. Accadia, F. Acernese, C. Adams, R. Adhikari, and et al., arXiv:1112.6005v4 (2012), arXiv:1112.6005 [astro-ph.CO].
[30] F. Beauville, M.-A. Bizouard, L. Blackburn, L. Bosi, L. Brocco, D. A. Brown, D. Buskulic, F. Cavalier, S. Chatterji, N. Christensen, A.-C. Clapson, S. Fairhurst, D. Grosjean, G. Guidi, P. Hello, S. Heng, M. Hewitson, E. Katsavounidis, S. Klimenko, M. Knight, A. Lazzarini, N. Leroy, F. Marion, J. Markowitz, C. Melachrinos, B. Mours, F. Ricci, A. Viceré, I. Yakushin, M. Zanolin, and joint LIGO/Virgo working Group, Classical and Quantum Gravity **25**, 045001 (2008), gr-qc/0701027.
[31] J. Aasi, J. Abadie, B. P. Abbott, R. Abbott, T. D. Abbott, M. Abernathy, T. Accadia, F. Acernese, C. Adams, T. Adams, and et al., Phys. Rev. D **88**, 062001 (2013), arXiv:1304.1775 [gr-qc].
[32] P. A. Evans, J. K. Fridriksson, N. Gehrels, J. Homan, J. P. Osborne, M. Siegel, A. Beardmore, P. Handbauer, J. Gelbord, J. A. Kennea, and et al., ApJS **203**, 28 (2012), arXiv:1205.1124 [astro-ph.HE].
[33] J. Aasi, J. Abadie, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, T. Accadia, F. Acernese, C. Adams, T. Adams, and et al., ApJS **211**, 7 (2014), arXiv:1310.2314 [astro-ph.IM].
[34] H.-S. Cho and C.-H. Lee, Classical and Quantum Gravity **31**, 235009 (2014), arXiv:1403.4681 [gr-qc].
[35] D. Nicholson and A. Vecchio, Phys. Rev. D **57**, 4588 (1998), gr-qc/9705064.
[36] M. Vallisneri, Phys. Rev. D **77**, 042001 (2008), gr-qc/0703086.
[37] M. Zanolin, S. Vitale, and N. Makris, Phys. Rev. D **81**, 124048 (2010), arXiv:0912.0065 [gr-qc].
[38] C. L. Rodriguez, B. Farr, W. M. Farr, and I. Mandel, Phys. Rev. D **88**, 084013 (2013), arXiv:1308.1397 [astro-ph.IM].
[39] W. E. East, S. T. McWilliams, J. Levin, and F. Pretorius, Phys. Rev. D **87**, 043004 (2013).
[40] E. A. Huerta and D. A. Brown, Phys. Rev. D **87**, 127501 (2013), arXiv:1301.1895 [gr-qc].
[41] E. A. Huerta, P. Kumar, S. T. McWilliams, R. O'Shaughnessy, and N. Yunes, Phys. Rev. D **90**, 084016 (2014), arXiv:1408.3406 [gr-qc].
[42] K. Belczynski, V. Kalogera, and T. Bulik, ApJ **572**, 407 (2002).
[43] P. C. Peters, Phys. Rev. **136**, B1224 (1964).
[44] T. Hinderer, B. D. Lackey, R. N. Lang, and J. S. Read, Phys. Rev. D **81**, 123016 (2010).
[45] I. W. Harry and S. Fairhurst, Phys. Rev. D **83**, 084002 (2011).
[46] W. G. Anderson, P. R. Brady, J. D. E. Creighton, and E. E. Flanagan, Phys. Rev. D **63**, 042003 (2001).
[47] R. Balasubramanian, B. S. Sathyaprakash, and S. V. Dhurandhar, Phys. Rev. D **53**, 3033 (1996), gr-qc/9508011.
[48] P. Ajith and S. Bose, Phys. Rev. D **79**, 084032 (2009), arXiv:0901.4936 [gr-qc].
[49] S. Fairhurst, J. Phys. Conf. Series **484**, 012007 (2014), arXiv:1205.6611 [gr-qc].
[50] T. Cokelaer, Phys. Rev. D **76**, 102004 (2007), arXiv:0706.4437 [gr-qc].
[51] I. W. Harry, B. Allen, and B. S. Sathyaprakash, Phys. Rev. D **80**, 104014 (2009), arXiv:0908.2090 [gr-qc].
[52] C. Van Den Broeck, D. A. Brown, T. Cokelaer, I. Harry, G. Jones, B. S. Sathyaprakash, H. Tagoshi, and H. Takahashi, Phys. Rev. D **80**, 024009 (2009), arXiv:0904.1715 [gr-qc].
[53] P. Ajith, N. Fotopoulos, S. Privitera, A. Neunzert, N. Mazumder, and A. J. Weinstein, Phys. Rev. D **89**, 084041 (2014), arXiv:1210.6666 [gr-qc].
[54] G. M. Manca and M. Vallisneri, Phys. Rev. D **81**, 024004 (2010), arXiv:0909.0563 [gr-qc].
[55] S. Privitera, S. R. P. Mohapatra, P. Ajith, K. Cannon, N. Fotopoulos, M. A. Frei, C. Hanna, A. J. Weinstein, and J. T. Whelan, Phys. Rev. D **89**, 024003 (2014), arXiv:1310.5633 [gr-qc].
[56] I. Mandel, C. P. L. Berry, F. Ohme, S. Fairhurst, and W. M. Farr, Classical and Quantum Gravity **31**, 155005 (2014), arXiv:1404.2382 [gr-qc].
[57] E. Barankin, Annals Math. Stat. , 477 (1949).
[58] R. McAulay and L. Seidman, IEEE Trans. Inf. Theory **15**, 273 (1969).
[59] R. McAulay and E. Hofstetter, IEEE Trans. Inf. Theory **17**, 669 (1971).
[60] L. P. Singer, *The needle in the hundred square degree haystack: The hunt for binary neutron star mergers with LIGO and Palomar Transient Factory*, Ph.D. thesis, California Institute of Technology (2015).
[61] C. P. L. Berry, I. Mandel, H. Middleton, L. P. Singer, A. L. Urban, A. Vecchio, S. Vitale, K. Cannon, B. Farr, W. M. Farr, P. B. Graff, C. Hanna, C.-J. Haster, S. Mohapatra, C. Pankow, L. R. Price, T. Sidery, and J. Veitch, ApJ **804**, 114 (2015), arXiv:1411.6934 [astro-ph.HE].
[62] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, eds., *NIST Handbook of Mathematical Functions* (Cambridge University Press, New York, NY, 2010) print companion to [63].

[63] DLMF, "NIST Digital Library of Mathematical Functions," http://dlmf.nist.gov/, Release 1.0.8 of 2014-04-25, online companion to [62].

[64] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, ApJ **622**, 759 (2005).

[65] K. Cannon, R. Cariou, A. Chapman, M. Crispin-Ortuzar, N. Fotopoulos, *et al.*, ApJ **748**, 136 (2012), arXiv:1107.2665 [astro-ph.IM].

[66] C. Pankow, P. Brady, E. Ochsner, and R. O'Shaughnessy, Phys. Rev. D **92**, 023002 (2015), arXiv:1502.04370 [gr-qc].

[67] M. Burgay, N. D'Amico, A. Possenti, R. N. Manchester, A. G. Lyne, B. C. Joshi, M. A. McLaughlin, M. Kramer, J. M. Sarkissian, F. Camilo, V. Kalogera, C. Kim, and D. R. Lorimer, Nature **426**, 531 (2003), astro-ph/0312071.

[68] D. A. Brown, I. Harry, A. Lundgren, and A. H. Nitz, Phys. Rev. D **86**, 084017 (2012), arXiv:1207.6406 [gr-qc].

[69] J. W. T. Hessels, S. M. Ransom, I. H. Stairs, P. C. C. Freire, V. M. Kaspi, and F. Camilo, Science **311**, 1901 (2006), astro-ph/0601337.

[70] T. P. Robitaille, E. J. Tollerud, P. Greenfield, M. Droettboom, E. Bray, T. Aldcroft, M. Davis, A. Ginsburg, A. M. Price-Whelan, W. E. Kerzendorf, A. Conley, N. Crighton, K. Barbary, D. Muna, H. Ferguson, F. Grollier, M. M. Parikh, P. H. Nair, H. M. Unther, C. Deil, J. Woillez, S. Conseil, R. Kramer, J. E. H. Turner, L. Singer, R. Fox, B. A. Weaver, V. Zabalza, Z. I. Edwards, K. Azalee Bostroem, D. J. Burke, A. R. Casey, S. M. Crawford, N. Dencheva, J. Ely, T. Jenness, K. Labrie, P. Lian Lim, F. Pierfederici, A. Pontzen, A. Ptak, B. Refsdal, M. Servillat, and O. Streicher, A&A **558**, A33 (2013).

[71] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical recipes," (Cambridge Univ. Press, 2007) Chap. 2.7.4, pp. 81–82, 3rd ed.